SCIENCE-CENTRIC SAMPLING APPROACHES OF GEO-PHYSICAL ENVIRONMENTS FOR REALISTIC ROBOT NAVIGATION

A Thesis
Presented to
The Academic Faculty

by

Lonnie T. Parker, IV

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2012

SCIENCE-CENTRIC SAMPLING APPROACHES OF GEO-PHYSICAL ENVIRONMENTS FOR REALISTIC ROBOT NAVIGATION

Approved by:

Professor Ayanna M. Howard, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Jeff Shamma
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Fumin Zhang
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Mike Stilman
College of Computing
Georgia Institute of Technology

Professor George Riley
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: June 15, 2012

DEDICATION

Based upon the infallible word of God,

I dedicate this work to Him, according to Proverbs 16:3.

ACKNOWLEDGEMENTS

Before anyone else, I want to acknowledge and thank my Lord and Savior, Jesus Christ, whom I am in and who lives within me. This degree and this experience would not be possible without His provision, favor, and grace. Especially in the moments when I did not believe it was possible to finish, my faith in Him pushed me forward. Next, to my wife, Andrea, who has been my constant encourager, supporter, and living reminder that I could complete this journey, thank you. Simply your presence in my life has made this walk all the more bearable. I love you. Next, to my parents, Lonnie III and Rosalia, thank you for your unwaivering support and consistent concern for me as I transitioned into adulthood during this degree. To my other family from across the country and around the way, Octavia, Vincent and Isabella "Izzy" Grimes, to the Bakers, and to the Beavers, thank you for welcoming me into your lives and for expanding my network of loved ones who have been praying to our God on my behalf. Thank you to Dr. Ayanna Howard for your support and commitment to me as an advisor and thank you to my thesis committee. To the funding organizations that have generously provided me with fellowships, scholarships, stipends and grants to support my research and fund my degree during my tenure at Georgia Tech: the Science, Mathematics, and Research for Transformation (SMART) Program, the Naval Undersea Warfare Center, the FACES program, the President's Office, and the National Aeronautics and Space Administration (NASA) under the Earth Science and Technology Office, Applied Information Systems Technology Program. To Dean Gary May, Dr. Comas Haynes, Dr. Raheem Beyah, the faculty and staff on the FACES Steering committee, the Office of Minority Educational Development, and the FOCUS program, thank you for helping to create a welcoming environment that

enables students like me to succeed at Georgia Tech. I am forever grateful for the experience. Lastly, I want to thank my brothers in Christ, Dr. Brad Jones, Dr. J. Chris Ford, Wendell C. Smith, and everyone from the Christian Organization of Dynamic Engineers (CODE). For your prayers in faith and for opening up your hearts to myself and one another, I thank you for your commitment to our Lord and I pray that your trust in Him will grow unbounded. Continue to exhibit not those qualities common to men of this world, but instead, those fruits of the Holy Spirit, common to men of God, love, joy, patience, peace, kindness, goodness, faithfulness, gentleness, and self-control.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The objective of this research effort is to provide a methodology for assessing the effectiveness of sampling techniques used to gather different types of geo-physical information by a robotic agent. We focus on assessing how well unique real-time sampling strategies acquire information that is, otherwise, too dangerous or costly to collect by human scientists. Traditional sampling strategies and informed search techniques provide the underlying structure for a navigating robotic surveyor whose goal is to collect samples that yield an accurate representation of the measured phenomena under realistic constraints. These sampling strategies are alternative improvements that provide greater information gain than current sampling technology allows. The contributions of this work include the following: 1) A method for estimating spatially distributed phenomena, using a partial sample set of information, that shows improvement over that of a more traditional estimation method. 2) A method for sampling this phenomena in the form of a navigation scheme for a mobile robotic survey system. 3) A method of ranking and comparing different navigation algorithms relative to one another based on performance (reconstruction error) and resource (distance) constraints. We introduce a specific class of navigation algorithms as example sampling strategies to demonstrate how our methodology allows different robot navigation options to be contrasted and the most practical strategy selected.

CHAPTER I

INTRODUCTION

Providing more accurate information about features of the Earth's surfaces, including physical phenomena, like elevation or mineral density, is a prime application for robotic surveying. Future federal mandates could require that Earth-observing systems (EOS) integrate data from more than just the remote sensing and static on-ground monitoring technologies that are currently employed. Excerpts from a recent 2010 investigation by committees of the National Research Council of the National Academies report there is a strong need for a national geodetic infrastructure [8]. With that need comes the challenge to identify which technologies will further the acquisition of more accurate geodetic information. According to the National Oceanic and Atmospheric Association (NOAA), the term *geodesy* is defined as "the science of measuring and monitoring the size and shape of the Earth and the location of points on its surface" [9]. NOAA and other organizations are responsible for the acquisition and improvement of geodetic information, both nationally and world-wide. More specifically, the National Academies study investigates "geodetic observing systems" in the larger context of a geodetic infrastructure and includes sea-level change monitoring, autonomous navigation, precision agriculture, civil surveying, earthquake monitoring, forest structural mapping, and biomass estimation as areas of application [8]. Each of these areas requires the capabilities that recent advancements in robotics offers. As such, in this thesis, we focus on measuring the effectiveness of algorithms used to acquire relevant geodetic and other types of information according to unique sampling techniques that are augmented by realistic robotic navigation methods.

One current challenge to measuring phenomena at different locations around the Earth is the task of obtaining highly detailed information at small spatial scales. While satellites equipped with high-performance sensors can provide a global assessment of the changes taking place at the surface of this planet, scientists still require in situ (*i.e.*, on the ground) validation of these measurements. Researchers in the geospatial community project the importance of this added perspective [10–13] while others in the robotics community propose solutions with varying degrees of success [14, 15]. A theory for how to systematically acquire this high-resolution information is best addressed through the task of robotic surveying and sampling based on the safety and flexibility benefits that robot-based advancements offer. We will discuss these benefits in greater detail in subsequent sections. A layperson's introduction to land surveying is publically available online, but in the context of robotics, [16] provides a useful overview by contrasting traditional surveying with humans versus using low-cost mobile sensors. The authors establish the usefulness of mobile autonomous agents to the Earth science community and show how these systems provide long-term value to projects outlined in the 2007 NASA Decadal survey [17]. Specifically, they discuss the utility of a robotic survey system (RSS) and its potential for augmenting information from already existing Earth-observing systems (EOS).

Whether surveying using multiple reference points or performing a sampling operation at a single location, we use the terms *surveying* and *sampling* interchangeably. The specific contributions of this research are listed below:

- Contribution 1: A method for estimating spatially distributed phenomena, using a partial sample set of information, that shows improvement over that of traditional, single-trial Bernoulli estimation.

- Contribution 2: A navigation method for sampling spatially-distributed phenomena, influenced by the informed estimation method of Contribution 1, that shows improvement over that of traditional navigation/sampling approaches.

- Contribution 3: A method of ranking and comparing sampling configurations relative to one another based on performance (reconstruction error) and physical (resource) constraints.

We employ different sampling strategies to acquire detailed information about an unknown area, estimate all remaining unvisited locations, and quantitatively measure each sampling strategy's success in meeting performance and resource constraints. The remainder of this thesis is organized as follows:

Chapter 2 A background of spatial estimation for environmental phenomena with a discussion of common techniques, terminology, and real-world applications.

Chapter 3 The theory of our informed estimation methodology, including the metrics necessary to successfully estimate spatially distributed data, and metrics for constrasting unique sampling configurations based on performance and resource constraints.

Chapter 4 The theory behind navigation methods used as sampling schemes.

Chapter 5 Results validating the theory of Chapters 3 and 4.

Chapter 6 Concluding remarks and recommendations for future work.

CHAPTER II

BACKGROUND & MOTIVATION

2.1 Traditional Spatial Sampling of Geo-physical Data

Sampling is a necessity in many fields, including the geo-physical sciences, due to the breadth of coverage required to estimate a particular phenomenon over a large area and the limited resources available to achieve that coverage. Examples of these phenomena are as follows: elevation in mountainous regions, chemical concentrations in soil, and mineral content or hazardous material levels in aquatic environments. Applications rarely exist where sampling is not preferred over the costly alternative of exhaustive coverage. Yet, determining how to collect a sufficient number of samples that will accurately represent a larger space is still an open problem and requires a more detailed scope, which we provide here.

One fundamental assumption made by all of the geoscience literature considered for this work is a priori access to the data set representing the entire sampling space via remote sensing technology [1,18–21]. Like most work in this genre, this assumption of unlimited data access is typically made because of the larger continental and regional scales at which such information is georeferenced [22]. Since remotely acquired imagery often serves as the primary source of the sample space to which sampling methods are applied, the a priori data access assumption fails to hold at the pixel scale where higher resolution information exists but is not accessible without in situ data collection. More information on this limitation is discussed in [16] and supports our interest in pixel-level detail (Figure 1).

Solving the problem of obtaining high-resolution detail beyond what remote imagery provides, regardless of the environmental phenomena, is a valuable asset to the

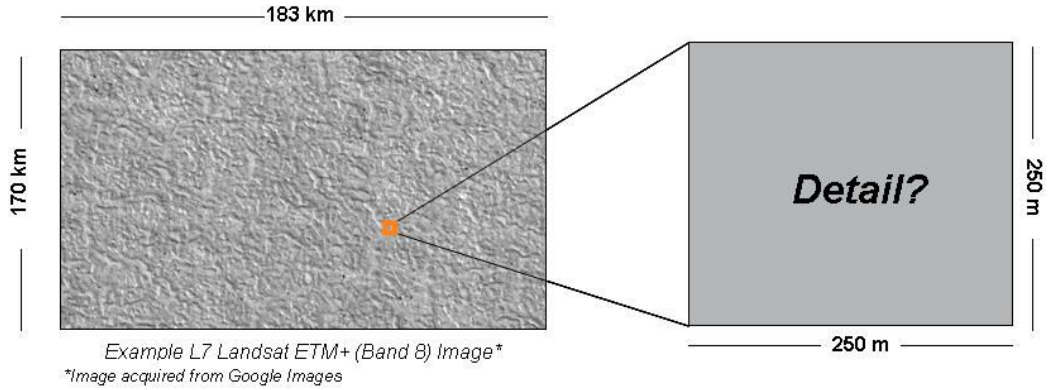


Figure 1: Extracted pixel-scale area of interest for a robotic survey system from a larger satellite image. Enlarged view of pixel provides no spatial variability information.

geoscience community and can be addressed with a robotic survey system (RSS). In our research, we thus focus our efforts on sampling an area with a robotic platform in such a way that we reduce reconstruction error and required resources. As with most robotic applications, we must integrate a plausible navigation strategy with our sampling approach such that it assesses some form of resource usage, *e.g.*, distance traveled, battery life, or time of experiment. Specifically, our methodology enables a scientist to contrast the effectiveness of different sampling schemes as a function of cost and resulting error from ground truth. In this work, we assume the cost of navigating to sampling locations to be greater than the cost of performing a sampling operation, which we discuss in more detail in Section 2.3. The issue we address in this thesis is how to evaluate different sampling strategies such that, with respect to a set of performance and resource constraints, the best strategy is selected.

The heterogeneity of an environment is particularly important when considering sampling methodologies as it can not be taken for granted that samples acquired locally in one portion of the sampling area are representative of the whole. We discuss heterogeneity of spatially distributed phenomena more in Section 2.3. The authors in [18, 23] cite heterogeneity as an important factor when considering how to sample an area to acquire an accurate spatial representation with minimal error. Additionally, [24] makes distinctions in how sampled data from the sampling area

may be represented. Some scientists will provide a categorical partitioning of the search space, where each partition contains a subcategory of a larger class, *e.g.*, grainy sand or rocky soil. Scientists may, instead, choose to report their findings in a quantized (or binary) format, *i.e.*, representing data as present or absent at a defined location. Surveys also exist that provide measurable values on a continuous scale at each location sampled to more fully capture the relative and absolute variability across the entire space [24]. We use sampling techniques that are inspired by robotic navigation literature, and metrics to describe the overall success of resulting sampling configurations.

Many of the approaches to sampling rely on variations of either evenly spaced or randomized selection schemes where samples are taken within stratified (or gridded) portions of the sampling space. Morrison references six sampling patterns (unaligned random, aligned random, unaligned stratified, aligned stratified, unaligned systematic, and unaligned systematic) designed to opimize sample placement in concert with sample quantity [1, 25] (Figure 2).

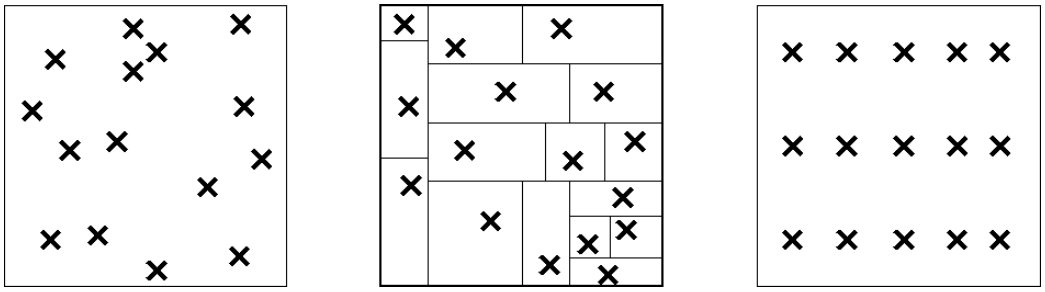


Figure 2: Three example sampling patterns (from left to right, unaligned random, unaligned stratified, and aligned systematic) used in [1] for optimizing sample placement in terrain measurment.

The history of these approaches to sampling is well established and widely applied, yet these schemes are usually considered primarily to attain improvement in computational efficiency, *i.e.*, a sampling strategy is selected that reduces time required to process observated data offline. The selection of sample points according to these schemes does not consider a policy or set of rules for navigating between each

discrete sample in real time. Additionally, these options rarely account for the cost of collecting all identified samples within the defined configurations, *i.e.*, the costs associated with sample-to-sample navigation. With these perspectives on traditional uses of sampling, there remains a disconnect between the task of sampling and the preferred methodology that ought to be employed by an agent (*i.e.*, human or robot) responsible for executing this sampling task in a real environment.

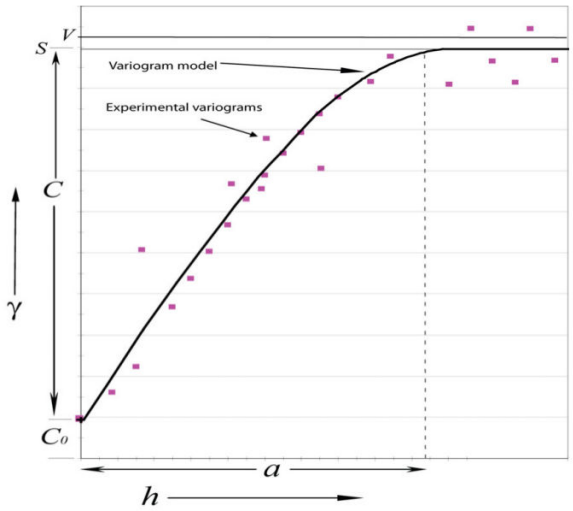
2.2 Reconstruction Through Interpolation

Once a set of samples is obtained, an accurate approximation of the space is made based on observed values in conjunction with the estimated values at unvisited locations. These estimated values are a function of the designated samples provided by a sampling scheme. An estimation process must be chosen that accounts for the proximity and frequency of observed samples relative to those that are not observed. The geostatistics literature provides methods of interpolation for extracting spatially-dependent information at unobserved locations based on those at observed points. There are many approaches to interpolation that tie together the importance of sample selection with resulting reconstruction error, a primary constraint in robotic surveying.

A comprehensive study of over twenty different interpolators (geostatistically-inspired and traditional) is available in [26]. This report from the Australian government contrasts over fifty relevant studies that use different interpolation options to improve specific error metrics. The study reveals the extreme subjectivity of how an interpolator should be chosen and the importance of identifying the correlation between a given interpolator and the spatial distribution of the sample set.

Estimation (or reconstruction) error, as it relates to interpolation, is also discussed in literature with respect to the spatial diversity exhibited by a particular phenomena and is often addressed by the geostatistics community in the form of the variogram.

This model-based approach to calculating the amount of spatial variation that exists across an area is a integral part of kriging, a widely accepted form of interpolation, particularly for data assumed to be generated by a random process [19, 23, 27]. As a model, the variogram has parameters that can be tuned, by hand or automatically, to approximate error with the lowest variance possible (Figure 3). The variogram,



$$\gamma = C_0 + C * \left(1.5 * \left(\frac{h}{a}\right) - 0.5 * \left(\frac{h}{a}\right)^3\right)$$

Figure 3: Example experimental variogram fit with a spherical model [2].

therefore, is a subjective estimation tool, calculated as a function of observations made a priori from the area of interest (Section 2.1). How these observations are made is, many times, based on a random selection of points, rather than from a deliberate sampling design process. In some cases, explicit consideration is given to the features that a sampling scheme should exhibit, *e.g.*, regular or irregular spacing [20], yet the number of parameters that influence a chosen variogram's performance prevent a direct tie between kriging and the spatial patterns of these sample designs. There exists a disconnect between how to choose a sample set that will yield the best (*i.e.*, lowest) reconstruction error across the space and the interpolation methodology used to generate that estimation error from that set. We address that disconnect and

make recommendations for metrics that quantify performance of potential sampling designs.

2.3 Environment Scope and Sampling Space

This research refers to "environmental phenomena" as any spatially distributed data set defined in a two-dimensional space. An important consideration in defining the search space is how the phenomena within the environment will be characterized, *i.e.*, determining the amount of spatial variability from one measurement location to another. The spatial relationship between data in a space is related to the concept of heterogeneity in the geospatial/sampling communities. Wang *et al.* conducts studies related to heterogeneous geographical data [18]. These surfaces are such that given a moving window of a size smaller than the area of interest, the statistical mean of the measurements within that window vary as a function of position and window size (Figure 4).

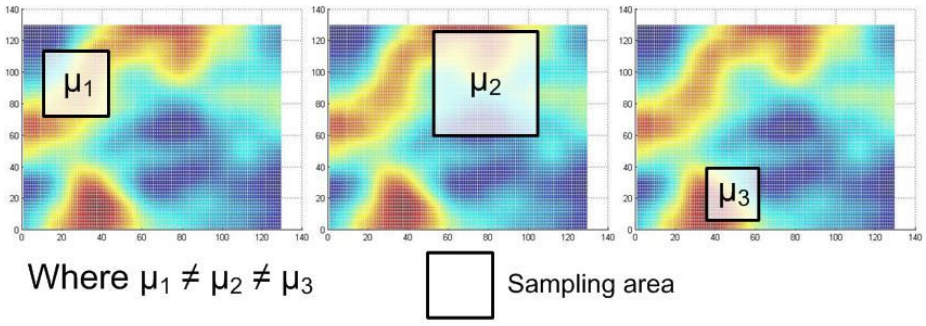


Figure 4: Pictorial representation of first-order heterogeneity within a 2D space.

The term "heterogeneity" is a useful descriptor for phenomena that varies spatially. Relevant spaces are those with phenomena changing over a period of months and/or years, *e.g.*, applications like ice sheet monitoring [12, 28, 29]. Characterizing changes in elevation for Arctic environments is especially useful as it allows better preparation of emergency response measures to the recent rise in sea levels in the past decade due to global warming [17]. Other examples include the distribution of soil

nutrients (*e.g.*, carbon, nitrogen, and phosphorus [30]) for mapping of soil properties and mineral deposits (*e.g.*, gold, iron, and lead). Related literature aims to shed light on the distribution of naturally-occuring spatial features common to specific phenomena. For example, understanding why gold is not commonly found evenly dispersed over a space, but instead is found in discrete amounts at spatially uncorrelated locations and in varying quantities motivates research in this area. The process of mapping these elements is widely investigated in the precision agriculture and mineral prospecting literature [22,31–33], while broader applications are discussed in [17]. Our work, specifically, is developed for monitoring more temporally-static heterogeneous phenomena, *i.e.*, information assumed to not change for a pre-defined time period. The theory behind our methodology is applicable to larger tasks discussed in literature that have gained attention on the national and international Earth science stage such as radiation sampling in hazardous environments and produce yield in precision agriculture [8].

2.3.1 Practical Considerations

In order to utilize a robotic survey system for sampling, the environment to be sampled must be navigable. This presumes the use of a platform that can successfully maneuver from one given location to another with a minimum payload capability. The payload, itself, should be equivalent to the weight of the sensor required to measure the phenomena of interest.

Given the range of variability that environmental phenomena can assume, there exist various trade-offs in realizable hardware that should be considered. For example, while the payload required to measure elevation may be inexpensive (*e.g.*, a mounted altimeter passively collecting data), navigating over rocky terrain can introduce an exhaustive amount of local spatial uncertainty in measurements. Therefore, an all-terrain platform with reliable inertial measurement equipment is preferable.

In contrast, the case of measuring the dispersion of a chemical spill from an isolated source over a relatively flat area poses a different set of constraints. While high-performance mechanical robustness of the platform may not be necessary, the challenge associated with transporting scenario-specific equipment and performing the sampling function is a relevant cost. A scenario that resembles the latter is discussed by Tunstel *et al.*. The authors reference a sensing system reminiscent of traditional surveyors. Although costs for sampling are not quantified directly as a function of the sensing instrument, the navigation algorithms discussed by Tunstel still cater to the technology's line-of-sight operation [34–36].

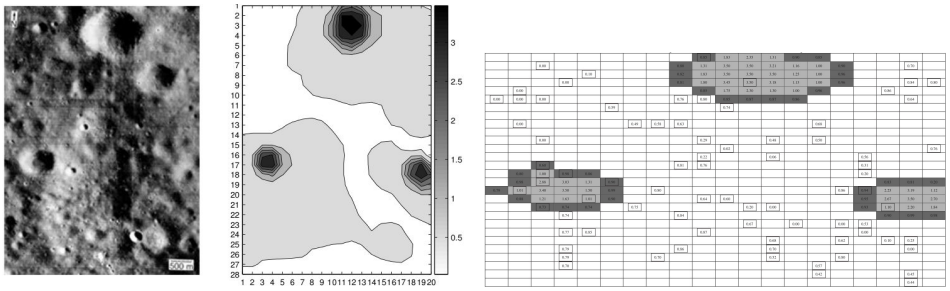
For a physically, realizable robotic hardware solution, the design of a robotic survey system must account for the environment to ensure realiable data is acquired consistently throughout the survey.

2.4 Previous Robotic Environmental Monitoring Systems and Applications

The next generation of in situ sampling technologies will be manifested in the form of robotic surveyors. Current technology that is used in land surveying procedures conducted by humans often comes in the form of immobile measurement equipment such as high performance lasers, electronic levels, and precision global positioning systems (GPS) [36]. In addition, useful, direct human management is required to successfully operate and reconfigure these devices, which can be tedious and exhausting. The next step of improvement suggests using mobile robotic units with the necessary measurement equipment to complete similar tasks as the human. This capability is particularly advantageous when conducting Earth science experiments, for example, when exploring Arctic terrain in harsh climates as a way to aleviate the dangers faced by human scientists [16]. Other applications extend to agriculture, foresty, construction site planning, and interplanetary exploration [34, 37–39].

K. H. Low *et al.* demonstrate work that addresses the sampling/coverage problem

in the context of robotics, discussing a coordinated effort between robotic agents to assess the distribution of measurable phenomena [4,40]. The contribution is two fold, as the authors first identify areas that contain "hot spots" or concentrated amounts of a particular type of sensor data (Figure 5). Then, they aim to reduce the bias that



(a) (b)

Figure 5: Spatial distribution of sampling locations within a gridded environment to locate "hot spots" of mineral content [3, 4].

inherently develops in approximated data using a series of popular statistical estimators. This work introduces their adaptive cluster sampling (ACS) method whereby each agent from a larger set selects surrounding grid locations to be included in its local network based upon whether or not the data at that location exceeds a predefined threshold. Other sampling strategies are discussed, including raster scanning, simple random sampling, and stratified random sampling. These sampling schemes are analogous to those discussed in [25] (Section 2.1). Although the types of tasks identified by the authors (*e.g.*, identifying mineral deposits and localization of chemical sources on other planets) are suitable for the ACS scheme, environmental phenomena is not always best characterized as discrete concentrations, as seen in elevation measurements that are spatially varied. Similarly, with respect to the estimation portion of their work, the aim is to minimize the statistical bias of the mineral concentration estimated, not at all focusing on the details of change of concentration between any two locations within the search space. When considering the sampling of terrain elevation, for example, this spatial change is important not only at the largest peak elevations, but in areas where smaller minima and maxima exist, therefore, ensuring sufficient

coverage is important. Correctly approximating the entire search space is relevant to the sampling task, not just where the phenomena's gradient or concentration is largest.

Rahimi *et al.* contributed additional robotic work specifically tackling the problem of environmental phenomena modeling in a real environment [5]. The aims of these authors' project, networked info-mechanical systems (NIMS), are more in line with our work, yet, their application is limited to the narrower domain of measurement within forestry and task scheduling. The authors apply a series of sampling strategies, much like those employed by [4, 40], to measure phenomena along an x-z or y-z plane (Figure 6). This is accomplished by discretizing the sample space and requiring

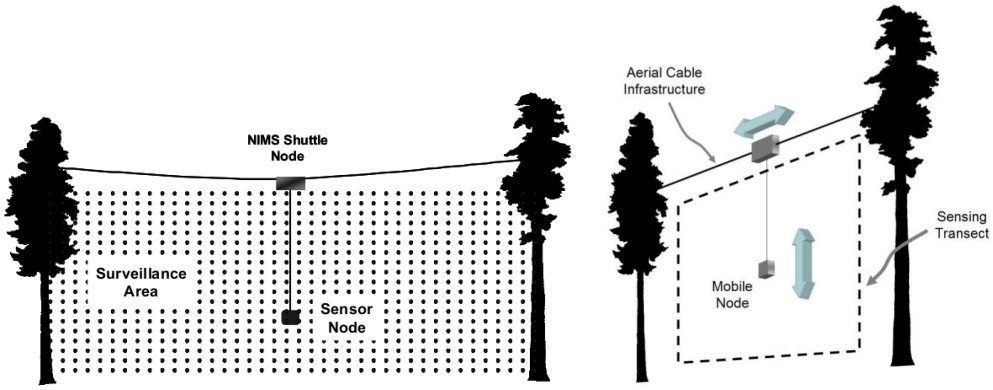


Figure 6: Forest-based experimental set-up to monitor environmental phenomena [5, 6].

repeat sampling at higher spatial resolutions where variance thresholds are exceeded inside specific "strata" or cells [5]. This sampling approach is termed nested stratified sampling. The navigation decisions of the mobile agent, however, are influenced by a network of static sensors embedded within the search space, effectively guiding subsequent sampling locations by the mobile agent. Using static sensors provides a priori data, an assumption we do not make in our work. The use of the variogram, a technique common to geostatistics, is also a ftool discussed by the authors. The variogram is included to emphasize the spatial variablilty of the data according to

their sampling strategy. Although the work meets a unique need for these types of forest-like environments, this is a very narrow scope of operation within the broader set of possible environmental-monitoring applications. Also missing from their work is a proposed measurement validation method.

In constrast to the previously mentioned work, other physical systems (tasked with monitoring environmental phenomena) have been tested by NASA Ames Research Center. The Haughton Crater field tests are a continuation of work from 2009 that incorporate two K10 plantary rovers equipped with high-profile science instruments [41]. These platforms produce high-resolution maps of the test area using light detection and ranging (lidar) and ground penetrating radar (GPR) technology primarily for the purpose of obstacle avoidance and path planning. Included on the science end of these missions is geologic mapping and geo-physical surveys for planetary exploration. The primary motivation of the NASA Ames work, however, is increased insight into the integration of both robotic and human teams. The actual robotic tasks for the K10 platforms are, therefore, limited to feasibility assessments. More information on the details of this work is in [38, 39]. The specific application of our work, *i.e.*, designing strategic sampling approaches for more accurate spatial estimation and measuring their effectiveness, is not an explicit objective of the NASA Ames work.

Precision agriculture is another category of work tied to physically implemented robotic systems. Anderson *et al.* offers a practical approach to increasing the potential crop yield of farmers via a mission execution and planning system (MEPS) which coordinates farming tasks. This system organizes task allocation and navigation for a set of autonomous agents with task-relevant farming equipment [37]. Their work highlights improvements in soil sampling and chemical spraying, yet their design is not centered around using the measurements of sampled (or sensed) information to drive the navigation decisions used to collect future samples. They discuss a combination

of navigation-related tasks that, following an initial survey of an area, a farmer can perform more detailed investigations:

> *Detection vehicles apply sweeping tasks to identify possible contamination then generate go-to tasks during mission execution for sampling and cleanup vehicles to perform further testing and treatment of suspected locations. Topological mapping requires a dynamic mix of the two tasks during mission execution. First, one or more vehicles execute broad sweeps to develop a "rough" map, which is then refined by recursively performing more detailed sweeps within areas of high variability.* [37]

Although resources used during the farming tasks are probably taken into consideration, there is no mention in the cited work of how management of these resources affects the operation of MEPS or any navigation-related decisions.

Along the lines of realistic navigation planning, Jin outlines methodologies relevant to precision farming in his PhD thesis [42]. Jin's work presumes access to 3D terrain data at the pixel level to successfully plan 2D and 3D surveys. Predominantly centered around traditional lawnmower patterns, Jin establishes detailed heading requirements of the autonomous farming agent by intelligently decomposing the field of interest. This work is accomplished under the assumption of access to the topography of the search space and the high-resolution information it includes. We will show in Chapter 4 how we mimick the lawnmower pattern while accomplishing our navigation without this prior knowledge.

Another application of the lawnmower pattern, discussed in [7], promotes the adaptation of the parameters definining the pattern (*i.e.*, swath width and linearity). This work by Bourgeois *et al.* simulates survey times for a Northeast coastal survey, segmenting coastal areas of interest into smaller polygonal-shaped areas. For each area to be surveyed by the team, three different variations of a traditional lawnmower pattern were executed by a surface vehicle equipped with a bathemetry system.

The authors of the simulation survey system introduce adaptive parallel (AP), linear regression (LR), and piecewise-linear (PL) line methods (Figure 7). The first "pass"

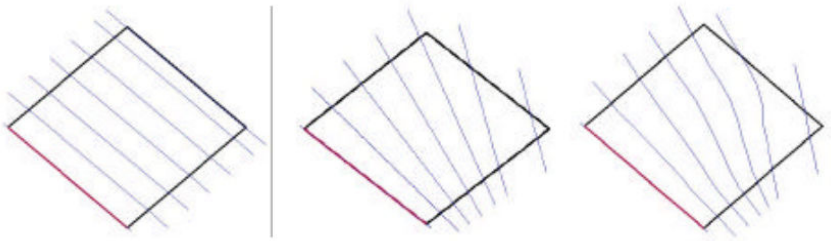


Figure 7: Tested navigation methods (AP, LR, and PL) based on linear transects, or "swaths" [7].

or swath for each method is defined at the boundary and every subsequent swath, thereafter, is determined by performing a fit to the previous line plus a shift to define the next set of waypoints for the vessel to follow. The outline of a particular swath is a function of its return signal strength as determined by the bathemetry system. The system assesses "the quality and coverage of the collected data and adaptively determines where the next survey line should be placed" [7]. The location of successive swaths within the sampling space do not adapt to the information gained from the measurements made online, but are instead adapting to the quality of return data of the coastal floor during execution of each swath. In our work, we contrast the difference between sample schemes that adapt to the data observed and those that do not. Additionally, we provide metrics for evaluating how well the sampling schemes tested meet pre-defined constraints (*i.e.*, error and resources) relative to one another.

Within the context of surveying, to our knowledge, the earliest work pioneering the use of robotics to explicitly perform surveying tasks was by Hashemi *et al.* [43,44]. Subsequently, his work was augmented by Tunstel *et al.* [34,35] in an effort to establish different navigation patterns useful for search and localization of gases and water on other planets. In the literature, a reigning theme is the assessment of navigation patterns using a quality of performance (QoP) metric. The QoP presumes that an

executed pattern will consist of continuous paths that are evenly and symmetrically distributed across the sampling space. Although useful for their specific work, the QoP metric is not designed to provide a valid quality value if these pattern placement conditions are not satisfied.

The common limitation found in the robotic-based research is that the environmental phenomena information observed is either ignored during navigation decisions or the information does not consider resource usage while seeking to obtain adequate coverage. As such, based on this previous work, we expand the potential of robotic surveying by offering metrics to quantify how sampling strategies, that meet practical, physical constraints, can generate the best reconstruction of the sampled space.

2.5 Broader Context: Mapping through Sampling

Many technological advances are used for mapping an environment. While our work is unique, it is advantageous at this point to clarify how our work is significantly different from other robot-based mapping applications.

Simultaneous localization and mapping (SLAM) is a well-documented approach for recording an agent's position within an environment while also extracting important features of that environment. First proposed as a robotics problem in the late 1980s [45], SLAM theory assesses geometric correlations between estimates of landmarks, relying on an increase in these correlations as the number of observations increase. The two goals of an agent executing a SLAM algorithm are: 1) to generate a map of the surrounding environment and 2) to generate a high-probability estimate of where that agent is physically located within that map. These goals are interdependent. While we seek to map the environment around us, we are not interested in tasks pertaining to localization of our robotic surveyor as we already presume access to that information via GPS. Also, we will not address obstacle awareness/avoidance

techniques, as SLAM typically considers. The environments we consider are unobstructed, whereas SLAM environments tend to include paths that require the presence of obstacles as landmarks, a necessity for the goals of correlation and position estimation.

Another initial comparison can be made between our work and that of wireless sensor networks (WSN). The many approaches available for deployment and maintenance of smaller sensors distributed across a sampled space make their employment a viable option. Even mobile sensors designed to re-configure throughout an area equivalent in size to the pixel footprint, as discussed in Section 2.1, are attractive alternatives [46,47]. A majority of the literature relevant to their use for environmental monitoring, however, is lacking in examples of practical hardware implementations. Of the simulated work found, the agents remain immobile and passively monitor the area of interest [48]. The coordination required to retrieve useful science data using multiple agents is a complex problem, therefore we actively seek out information about the sample space using a single agent. Where our work does share similarities with WSNs is in the usefulness of our algorithms to individual mobile nodes operating within a network. Under the conditions that each node is tasked with covering a subset of area within the sample space, application of our work by these nodes would collectively achieve the goal of sampling the entire area.

Shared between SLAM, WSN, and our work are underlying principles of statistical learning theory (SLT). The best description common to these approaches to mapping is how a function created by natural processes, f, is observed and recorded by another function, $\hat{f}$, whose goal is to estimate the expected difference to some degree of accuracy less than ϵ (Equation (1)).

$$E[|\hat{f} - f|] < \epsilon. \tag{1}$$

Although there exists partial overlap between these communities of research (Figure 8), we address the mapping problem for the Earth science community in ways

other approaches to mapping (*e.g.*, SLAM and WSNs) do not. We provide a solution tested in theory, software and hardware that prioritizes the sampling of a predefined area achieved within user-defined error and resource constraints by a single, manageable agent.

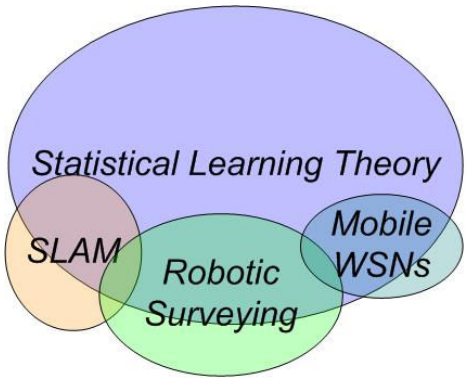


Figure 8: Intersecting scope of other robotic mapping applications.

CHAPTER III

METHODOLOGY: METRICS FOR SAMPLING SYSTEM ANALYSIS

In this chapter, we quantify the methods of estimating spatially-distributed phenomena, which assist in providing information on how specific sampling configurations should be chosen. These methods require defining a space (or environment) where samples (or observations) are collected, rules for estimating values at unsampled locations within that space, a ranking system to compare multiple sampling configurations, and additional constraints that further justify selecting a particular configuration.

3.1 Spatial Environment and Specifications

For our work, we estimate environmental phenomena by combining navigation algorithms, for sample selection, with interpolation techniques. We measure the effectiveness of these algorithms with respect to a set of error and resource constraints. These terrain spaces are initially modeled mathematically as a family of real-valued functions, Z, defined in $\mathbb{R}^2$. We say that for a given navigation algorithm, A, and a function, z from Z, $A(z)$ maps to a space of B discrete observations. A function, $\hat{z}_A$, is generated from B as an approximation of z, and has reconstruction error in the range of [0, $Error_{Max}$]. Part of our aim is to choose a navigation algorithm, A, whose trajectory maps to these B observations (or measurements) of z such that a function, $\hat{z}_A$, generated from those measurements will approximate z within a specific range of error. We discretize the sample space to more efficiently assess error between ground truth, z, and our approximated function, $\hat{z}_A$.

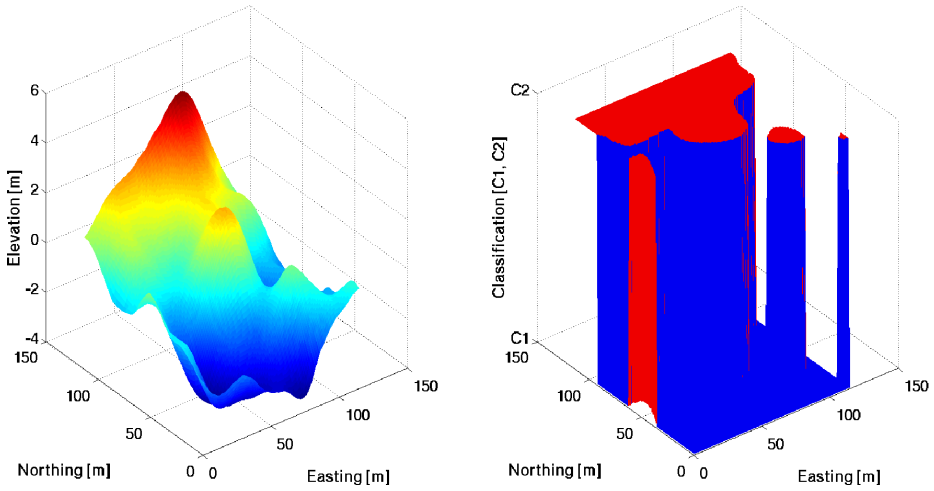
3.1.1 Discretization and quantization

To properly discretize our environment, let each location in the sample space be defined by the matrix $S = X_{dim} \times Y_{dim}$, where $X_{dim} = [0, \ldots, \delta, \ldots, \delta(M-1)]$, and $Y_{dim} = [0, \ldots, \delta, \ldots, \delta(N-1)]$. Here, M and N represent the quantity of rows and columns of samples along the x and y dimensions, respectively, and δ represents a scientist-defined spatial resolution, *i.e.*, the in-between sample spacing, such that, δ belongs to the set of all real numbers, $\mathbb{R}$. This discretization of the continuous space is equivalent to overlaying a grid on top of z and only accounting for data at intersecting grid locations. The discretization is also necessary for enabling scientists to define the desired level of lateral spatial accuracy within the pixel footprint (Section 2.1). This accuracy increases particularly as δ approaches 0. We define an upper limit, MN, as the total possible measurements (or observations) that can be collected by a robotic survey system. Additionally, for all x_i belonging to X_{dim} and for all y_j belonging to Y_{dim}, there exists a sample from the terrain, $z(s_{i,j})$. Let $s_{i,j} = (x_i, y_j)$ for $i \in 1 : M$ and $j \in 1 : N$.

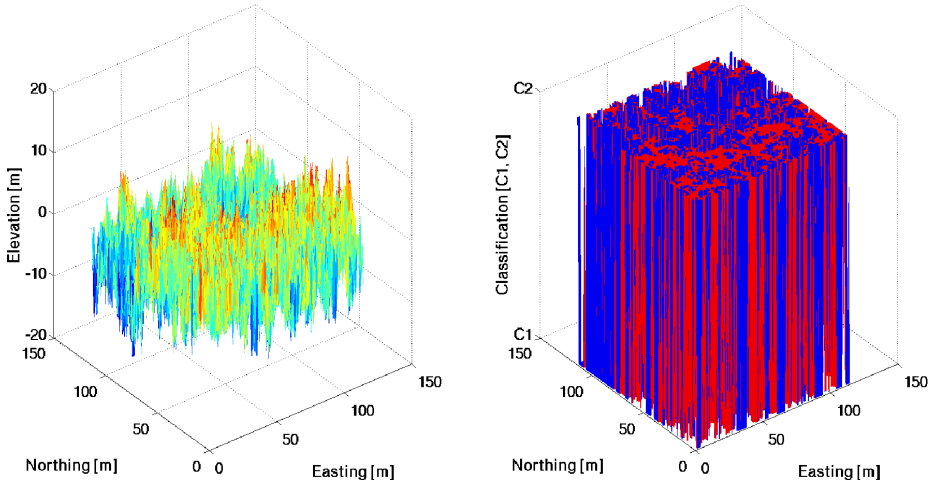
To properly define the nature of our measured phenomena, we perform a quantization from a continuous to dual-class scale of each sample value collected and for the ground-truth function, z. We let C represent a dual-class system defining the values of samples within the search space, S, where $C = [c_1, c_2]$. This quantization allows us to redefine z as f. We calculate the mean of all values in z at each sample location, $s_{i,j}$, and use the statistical mean, $\bar{z}$, as a threshold to define f (Equation (2)).

$$f = \begin{cases} c_1 & \text{if } z > \bar{z} \\ c_2 & \text{if } z \leq \bar{z} \end{cases} \tag{2}$$

This definition in Equation (2) generates a quantized family of functions, F. Subsequently, all samples collected by navigation algorithm A are also classified as belonging to one of two classes, c_1 or c_2, and thus the original approximation, $\hat{z}_A$, is hereafter redefined as $\hat{f}_A$ (Figure 9).



(a) Quantization of function z into function f for slowly varying values. Left plot: Standard deviation=1.845. Right plot: λ_{Avg}=1.5%.



(b) Quantization of function z into function f for erratically changing values. Left plot: Standard deviation=4.522. Right plot: λ_{Avg}=0.285.

Figure 9: Contrasting the quantization of terrains with different features and the analogy between standard deviation with λ_{Avg}.

Since obtaining reasonable reconstruction error is, in part, a function of sample placement, the spatial arrangement and correlation between the sampled values are an important attribute of the environment. As such, we define another condition under which our navigation algorithms are tested. Just as the standard deviation of values in a continuous function, z, range from small (*e.g.*, slowly varying elevation measurements) to large (*e.g.*, erratically changing soil moisture measurements), the same spatial frequency attribute is observed in a dual-class system, f. Specifically,

we sort example functions from class F into different types based on a metric of average spatial frequency, λ_{Avg}. This sorting is accomplished using information about the tendancy of a terrain's adjacent sample values to switch from one value to the other (*i.e.*, from c_1 to c_2, and vice versa) over the dimensions of the physical space. The value, λ_{Avg}, allows us to establish different types of terrain spaces and more readily identify which navigation algorithms are better suited for certain terrain types. Given that our search space is rectangular in nature, we calculate λ_{Avg} as the average number of changes between adjacent sample locations along the positive x-direction and positive y-direction for all rows and columns in the discretized space, respectively. The spatial frequency of a given terrain is defined as follows:

$$\lambda_{Avg} = 100\frac{\sum_{m_1=1}^{M}\sum_{n_1=2}^{N}\lambda^1_{m_1,n_1} + \sum_{n_2=1}^{N}\sum_{m_2=2}^{M}\lambda^2_{m_2,n_2}}{MN(M-1)(N-1)} \tag{3}$$

$$\lambda^1_{m_1,n_1} = \begin{cases} 1 & \text{if } [f(s_{m_1,n_1}) \neq f(s_{m_1,n_1-1})] : true \\ 0 & Otherwise \end{cases} \tag{4}$$

$$\lambda^2_{m_2,n_2} = \begin{cases} 1 & \text{if } [f(s_{m_2,n_2}) \neq f(s_{(m_2-1),(n_2)})] : true \\ 0 & Otherwise \end{cases} \tag{5}$$

In Equation (3), $\lambda^1_{m_1,n_1}$ is equal to "1" if a change occurs between positions s_{m_1,n_1} and $s_{m_1,(n_1-1)}$. Similarly, $\lambda^2_{m_2,n_2}$ is equal to "1" if a change occurs between locations s_{m_2,n_2} and $s_{(m_2-1),n_2}$ in the search space, S. In the event that no change occurs between adjacent positions, a value of "0" is assigned. Based on Equations (3)-(5), a visual depiction of how spatial frequency is calculated for a terrain is shown in Figure 10. In the example from Figure 10, there exists an average switching between classes detected in the function, f, of 6.5 (the average of six and seven when moving in the positive y and x directions, respectively). Identifying types of

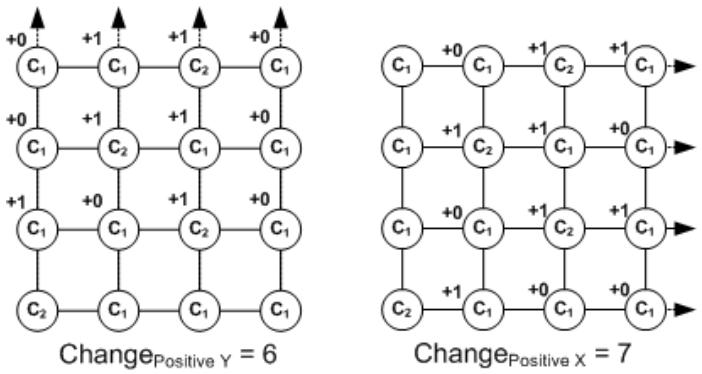


Figure 10: Frequency assessment of an example function, f.

terrains from F based on λ_{Avg} is analogous to sorting different types of continuously-valued environmental phenomena based on its spatial standard deviation. Separating smoothly varying terrain elevation values (Figure 9a) from erratically varying soil moisture measurements (Figure 9b) is an example of distinguishing between different types of data in a continuous domain. We confirm the relationship between standard deviation of terrains in the continuous space and λ_{Avg} of those terrains' quantized equivalent (Figure 11). In the next section, we will use λ_{Avg} to help us identify the

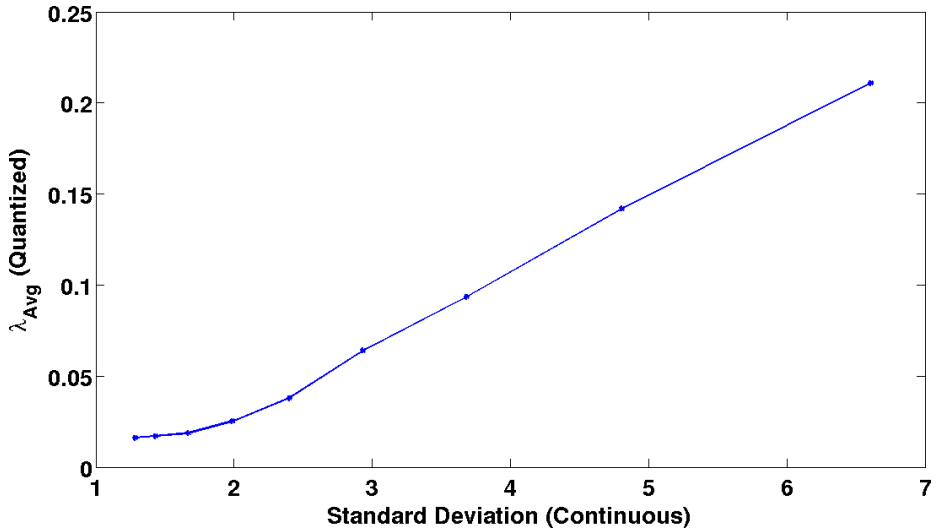


Figure 11: The linear relationship between standard deviation and λ_{Avg}.

quantifiable differences in performance between our estimation methodology and a traditional estimation method. Specifically, we can distinguish how well our method estimates spaces belonging to one spatial frequency, λ_{Avg}, versus spaces belonging to

another.

3.2 Classification-based Interpolation

Evaluating the performance of relevant navigation algorithms for robotic surveying depends heavily on the posteriori application of suitable interpolation methods to estimate information at unvisited, and therefore unsampled, locations. This dependence exists because of several important factors that include: where samples are collected, their observed value, and how they are related to one another, spatially, such that an accurate estimate can be made at these unvisited locations. It is important that the relationship between these factors appropriately compliments the theory of the particular interpolation technique used to estimate the unsampled locations in the search space.

3.2.1 Estimation and Probability

To better relate the theory behind our methodology of estimating unsampled locations, we consider the task of sampling and estimation as a nearest-neighbor classification problem. Based on the search space and the total number of possible observations, B, we outline the theoretical performance that a robotic surveyor can achieve in terms of probability of error generated based on its collected samples. For any number of samples, B, collected in the search space, S, let Q_B equal the number of unique combinations of B samples from all MN possible samples or observations (Equation (6)).

$$Q_B = \frac{MN!}{B!(MN-B)!} \tag{6}$$

For each unique sample configuration, q_h, where $h \in 1 : Q_B$, there exists $MN - B$ unobserved locations, in S, that require an estimation based on those B observed locations. In this chapter, we will discuss the performance of our methodology based on the total number of possible configurations, Q_{Total}, for values of MN equal to 9

and 16 (Equation (7)).

$$Q_{Total} = \sum_{v=1}^{MN} Q_v \tag{7}$$

It is important to note that Q_{Total} grows and shrinks parabolically when B is allowed to vary from a single observation to the total number of possible observations, MN, in S (Figure 12). The large increase in the scale of Q_{Total} from MN equal to 9 to MN

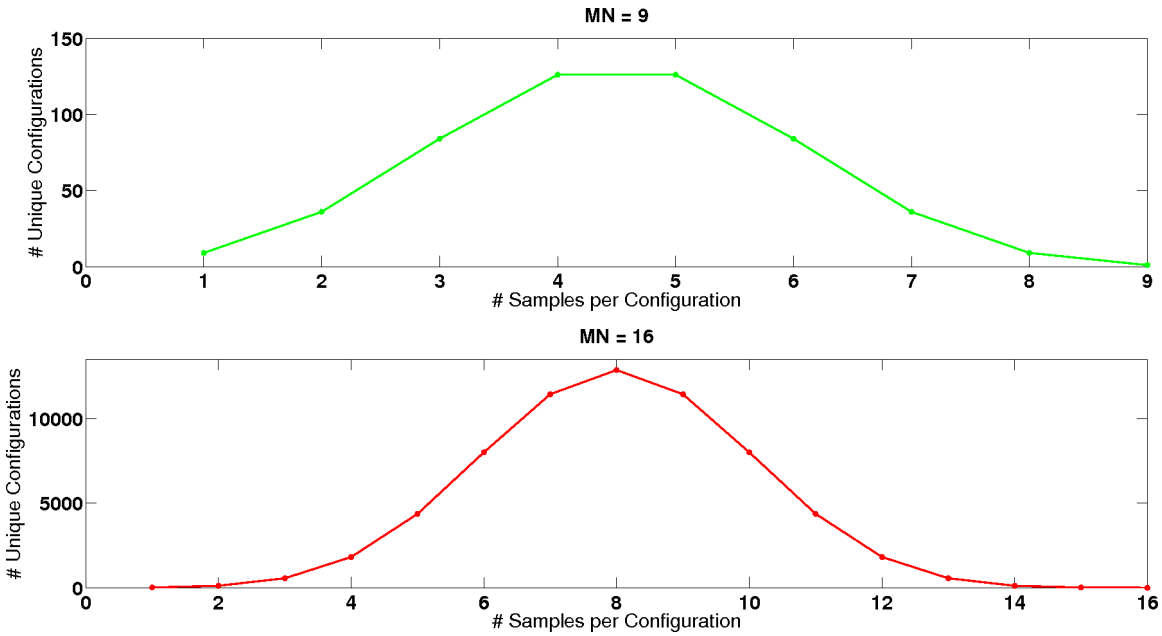


Figure 12: The trend of how the number of possible configurations in a space, S, changes as the number of samples in each configuration increases from the smallest to the largest number of possible observations, MN.

equal to 16 emphasizes the computational challenge in considering the performance of all possible sample configurations for larger values of MN.

We establish an estimation rule, from which we derive a theory of average expected error. Our estimation rule is based on a L-point nearest neighbor principle similar to kNN classification [49], where, given an unobserved location, $\hat{s}_{i,j}$, information about the quantity of observed locations, belonging to a particular class in C (Section 3.1), is combined with information of the proximity of those same samples to provide an estimated value at $\hat{s}_{i,j}$. The metrics drawn from our estimation process lead to an expected error for each estimate and, consequently, a total measure of expected error

for each configuration, q_h.

Let s, as a subset of S, be the set of B observations made within S, and let $\hat{s}$, as a subset of S, be the set of $MN - B$ unobserved locations at which values are approximated based on the values at s such that $(s \cup \hat{s}) = S$. Let $U = [u_1, u_2]$, where u_1 and u_2 are the quantity of nearest neighbor samples in set s, that belong to either class c_1 or c_2, respectively, and let $u_1 + u_2 = L$. Here, L is the maximum number of nearest neighbors used to estimate a value at a location.

For a given estimation location, $\hat{s}_{i,j}$, we also define R as the maximum lateral distance away from the estimate, in $\mathbb{R}^2$, where a nearest neighboring sample (from set s) may exist. R is a normalization factor calculated based on the maximum possible distance away from the estimate along X_{dim} and Y_{dim}, $R = [M + N - 1]$. R is used in our estimation rule defined in Equations (8)-(10).

$$\hat{f}_A(\hat{s}_{i,j}) = \begin{cases} c_1 & \text{if } \sigma_{c_1} > \sigma_{c_2} \\ c_2 & \text{if } \sigma_{c_1} < \sigma_{c_2} \\ w(\rho) & Otherwise \end{cases} \tag{8}$$

where $\hat{f}_A(\hat{s}_{i,j})$ is the estimate at unobserved location $\hat{s}_{i,j}$ based on samples collected by a navigation algorithm, A. Note that the value of the estimated function, $\hat{f}_A$, is equivalent to the value of the original function, f, at all visited locations, s, *i.e.*, $\hat{f}_A(s_{i,j}) = f(s_{i,j})$, for all s. The variable, ρ, is a uniform random variable defined between 0 and 1 and we define $w(\rho)$ as a function whose output is c_1 if $\rho < \gamma$ and c_2 if $\rho \geq \gamma$. The range for γ is between 0 and 1, although we assume a value for γ of 0.5 for all of our work. The values σ_{c_1} and σ_{c_2} are defined as follows:

$$\sigma_{c_1} = \frac{u_1}{L}[1 - \frac{d_{c_{1,Avg}}}{R}] \tag{9}$$

$$\sigma_{c_2} = \frac{u_2}{L}[1 - \frac{d_{c_{2,Avg}}}{R}] \tag{10}$$

In Equations (9)-(10), $d_{c_{1,Avg}}$ and $d_{c_{2,Avg}}$ represent the average distance between the estimate location, $\hat{s}_{i,j}$, and all nearest neighbors, in s, belonging to classes c_1 and c_2, respectively. The values represented by u_1 and u_2 weight the estimate, $\hat{f}_A(\hat{s}_{i,j})$, towards the more likely classification. Our expected error for sample configuration, $E[error_{q_h}]$, is a function of the size of the sample space, S, the unique spatial placement of each collected sample, $s_{i,j} \in s$, and the value (or classification) of that sample, $f(s_{i,j})$. Since we have a dual-class space, the expected error generated by q_h is stated as follows:

$$E[error_{q_h}] = \sum_{g=1}^{MN-B} \epsilon_{\hat{f}_A(\hat{s})_g} \tag{11}$$

Where we define $\epsilon_{\hat{f}_A(\hat{s})}$ for all $MN - B$ unvisited locations in the configuration, q_h, at a particular location in S as follows:

$$\epsilon_{\hat{f}_A(\hat{s})} = \gamma(1 - \eta) \tag{12}$$

In Equation (12), γ equals a constant value determined a priori and $\epsilon_{\hat{f}_A(\hat{s})}$ represents a probability of error for each estimated location, *i.e.*, a measure of likelihood that the g^{th} estimate, $\hat{f}_A(\hat{s})_g$, is incorrectly classified based on its nearest neighbor information. We calculate this error based on the confidence, η, in an unknown sample's estimated value. The confidence is defined in Equation (13).

$$\eta = |\sigma_{c_1} - \sigma_{c_2}| \tag{13}$$

For all our work, we assume γ equals 0.5 to assign equal likelihood that an estimated location will be classified as c_1 or c_2 in the event that η equals 0. We incorporate information of neighboring samples in our definition of probability of error, $\epsilon_{\hat{f}_A(\hat{s})}$, to acknowledge that as the quality of known information decreases, *i.e.*, a smaller quantity and greater distance away of neighboring samples, the expected error for each estimated location increases.

Actual error between the estimated function, $\hat{f}_A$, and the original function, f, is calculated as a sum of all estimated locations $MN - B$ that are incorrectly classified by our estimator (Equation (8)) and is defined in Equations (14)-(15).

$$error_{q_h}^{Actual} = \frac{\sum_{g=1}^{MN-B} \epsilon_g}{MN - B} \tag{14}$$

$$\epsilon_g = \begin{cases} 0 & \text{if } \hat{f}_A(\hat{s})_g = f(\hat{s})_g \\ 1 & \text{if } \hat{f}_A(\hat{s})_g \neq f(\hat{s})_g \end{cases} \tag{15}$$

In Equation (15), $\hat{f}_A(\hat{s}_g)$ is the estimated value at the g^{th} location, $\hat{s}_g$, while $f(\hat{s}_g)$ is the true value of the originally estimated function at that same location for B samples collected according to A.

Our definitions in Equations (8)-(15) enable us to generate an informed measure of expected and actual error for any dimension of space, S. Depending on the spatial frequency of f, these error metrics quantify how our informed estimation method outperforms the more traditional method of single-trial Bernoulli estimation [50].

3.2.2 Improvement over traditional estimation

We select single-trial Bernoulli estimation as a baseline method to compare the performance of our informed estimation method. We select single-trial Bernoulli estimation because of its relevance in generating a classification between one of two values. Single-trial Bernoulli estimation is typically employed to determine the probability of a "success" or "failure" for a particular event, *i.e.*, the probability, γ, that a flipped coin will land heads is considered a single-trial Bernoulli event. In our case, we consider the estimation of every unvisited location within the search space as an "event" and we use single-trial Bernoulli estimation to determine the value (or classification) at all $MN - B$ locations. Just as we defined our informed estimation rule, we define analogous metrics for this baseline method (Equations (16)-(18)).

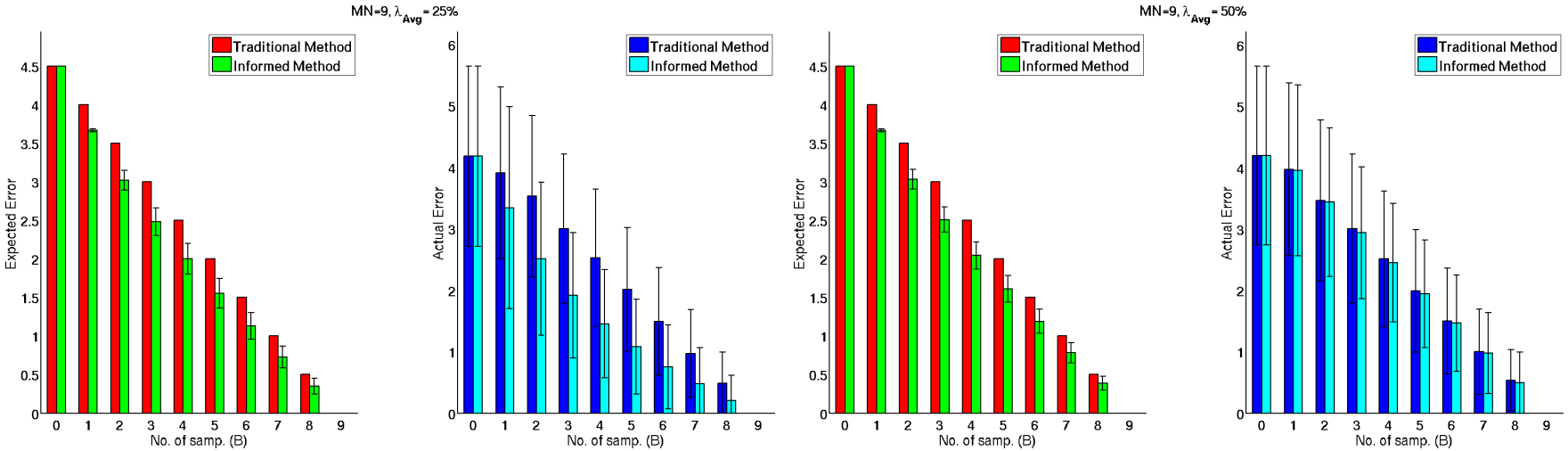
$$\hat{f}_A^{Traditional}(\hat{s}_{i,j}) = \begin{cases} c_1 & \text{if } \rho > \gamma \\ c_2 & \text{if } \rho \leq \gamma \end{cases} \tag{16}$$

$$E[error_{q_h}^{Traditional}] = \sum_{g=1}^{MN-B} \epsilon_{\hat{f}_A^{Traditional}(\hat{s})_g} \tag{17}$$

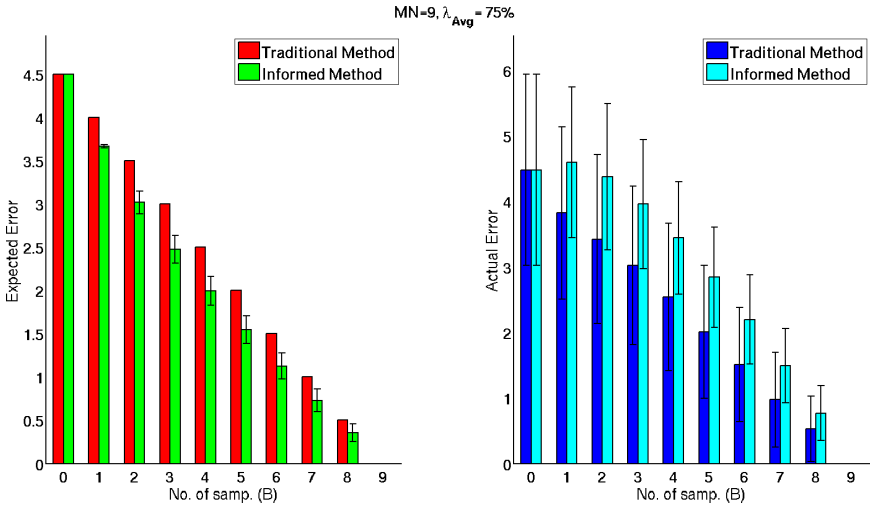
$$\epsilon_{\hat{f}_A^{Traditional}(\hat{s})} = \gamma \tag{18}$$

Again, γ may be defined a priori as a value between 0 and 1. We assume a value of 0.5 for our work, thus, the expected error generated according to single-trial Bernoulli estimation, $E[error_{q_h}^{Traditional}]$, is equal to a fraction of all $MN-B$ unsampled locations, $\gamma(MN-B)$. It should also be noted that the definition of the actual error generated according to single-trial Bernoulli estimation is identical to that for informed estimation from Equations (14)-(15), except $\hat{f}_A^{Traditional}(\hat{s})_g$ is used instead of $\hat{f}_A(\hat{s})_g$.

To verify the performance improvement of our estimation methodology over single-trial Bernoulli estimation (*i.e.*, "coin-flip" estimation), we evaluate the average expected and actual error generated by all sample configurations of a square space with MN samples and a specific spatial frequency of λ_{Avg}. In the first scenario, we define $M = 3$ and $N = 3$ to establish the discretized space, S, and consider all possible sample configurations consisting of B samples, where B is increased from 0 to $MN = 9$, $\lambda_{Avg} = 25\%$, and $L = 4$ (Figure 13a). In these graphs, we compare the output from our expected error calculations with the single-trial Bernoulli estimation method as the number of sampled locations, B, increases. We also compare these two methods as the spatial frequency of the terrain, λ_{Avg}, increases.



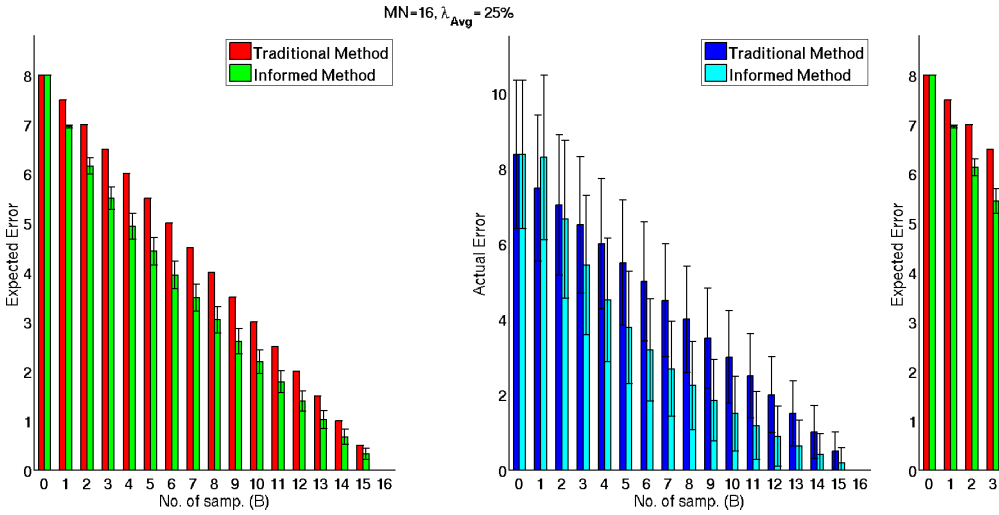
(a) $\lambda_{Avg} = 25\%$. (b) $\lambda_{Avg} = 50\%$.



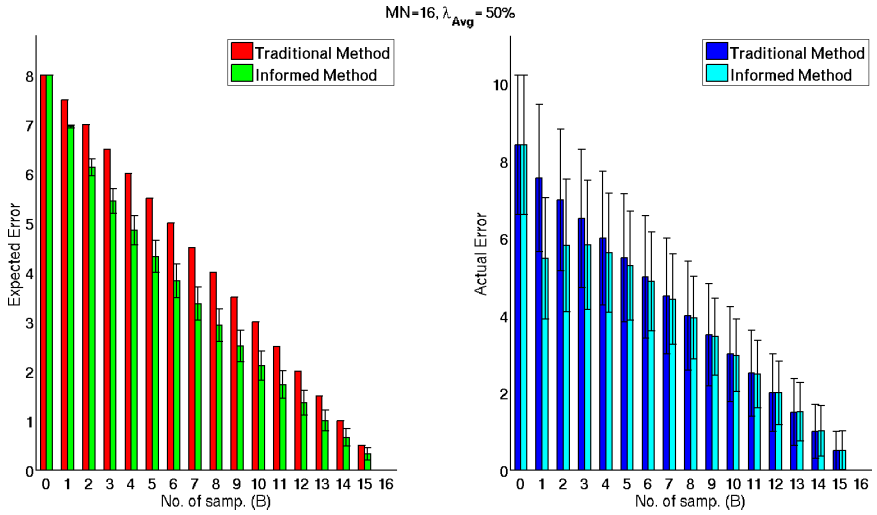
(c) $\lambda_{Avg} = 75\%$.

Figure 13: Comparing expected and actual error between traditional and informed estimation methods, $MN = 9$, $\lambda_{Avg} = 25, 50$, and $75\% \pm 8.3$.

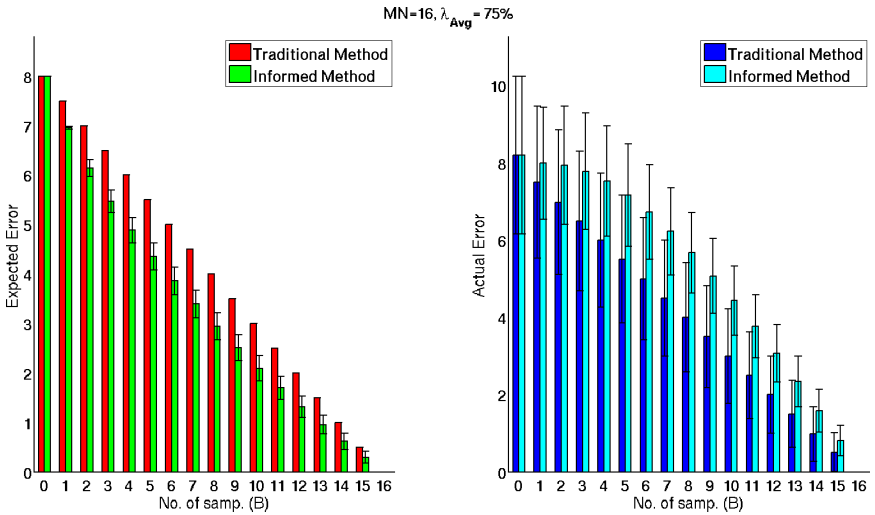
For all scenarios considered, the average expected error, $E[error_{q_h}]$, generated by our informed methodology outperforms the average expected error of the single-trial Bernoulli estimation method. Similarly, our method produces favorable average actual error when contrasted with that of the Bernoulli-based estimation (Figure 13). When the value of λ_{Avg} is increased to 50 percent, we notice that the actual error of our method begins to match the performance of the traditional method (Figure 13b). Finally, for $\lambda_{Avg} = 75\%$, we see a degredation in the performance of our informed method (Figure 13c). At this point, for estimation of any spaces exhibiting a spatial frequency greater than 50 percent, we default to using traditional, *i.e.*, single-trial Bernoulli, estimation. This test is repeated for $M = 4$ and $N = 4$ and B is increased from 0 to $MN = 16$ (Figure 14).



(a) $\lambda_{Avg} = 25\%$.



(b) $\lambda_{Avg} = 50\%$.



(c) $\lambda_{Avg} = 75\%$.

Figure 14: Comparing expected and actual error between informed and traditional estimation methods, $MN = 16$, $\lambda_{Avg} = 25, 50$, and 75%.

For each of the spatial frequencies considered (λ_{Avg} = 25, 50, and 75%) in our $MN = [9, 16]$ scenarios, we make three important observations: 1) As λ_{Avg} is increased (*i.e.*, $\lambda_{Avg} > 25$ %), based on the actual error calculations, we default to using the traditional, single-trial Bernoulli estimation technique. 2) The expected error generated for all possible configurations of B samples is not impacted significantly by data sampled at one spatial frequency versus another (left plots of Figures 13a-14c). 3) Of the terrain types where informed estimation shows the greatest improvement over traditional methods for actual data ($\lambda_{Avg} \leq 25$ %), the number of samples recommended per configuration is approximately 30-40 percent of the total possible number of samples in S.

3.3 Configuration Ranking: Expected and Actual Error

We observe, from the theoretical results in Section 3.2.2, that the number of samples chosen to estimate a space is important, but also as important is the unique spatial arrangement (or configuration) of those samples relative to one another and relative to the spatial frequency, λ_{Avg}, of the sample space, S. Suppose only B samples may be collected by a robotic survey system. This limitation, in terms of sample quantity, reduces the number of possible sample configurations considered for S from Q_{Total} (Equation (7)) to Q_B (Section 3.2.1). Still, depending on the size of B and MN, that amount, Q_B, can be quite large as shown in Figure 12. This is particularly true if B is approximately 30 percent of MN (Section 3.2.2). Once the upper limit of possible sample configurations is determined (based on B samples), it is necessary to identify which of the Q_B configurations will provide the best (*i.e.*, lowest) total expected error when applied to the space, S. Before we can discuss how a robot physically navigates between sample locations, we must determine which configuration of B samples provides greater benefit when ranked relative to alternate sample configurations of the same sample quantity, *i.e.*, B. This ranking is achieved by assessing the average

configuration confidence, η_{Avg}, for each configuration from all Q_B configurations possible (Figure 15). Our ranking metric, η_{Avg}, for each configuration of B samples, is

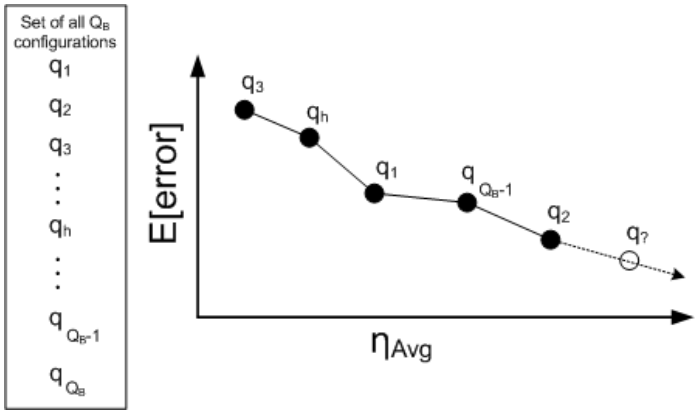


Figure 15: Ranking of all Q_B configurations according to confidence measure, η_{Avg}.

the average measure of confidence that our informed estimation method has correctly estimated each of the remaining $MN - B$ unobserved locations (Equation (19)).

$$\eta_{Avg} = \frac{\sum_{i=1}^{MN-B} \eta_i}{MN - B} \tag{19}$$

In Equation (19), η_i represents the measure of confidence (or certainty) that the estimation of the i^{th} unsampled location is accurate. This ranking metric is calculated given a particular sampling configuration, q_h, for $h \in 1 : Q_B$. The average of these confidence values, η_i, measured across all unsampled locations ($MN-B$), therefore, is represented as η_{Avg}. This measure of η_{Avg} is important because, from it, we can project the total expected error that a specific sampling configuration will generate based on a confidence measure of how well we believe its remaining unsampled locations will be estimated.

To show the relevance of this ranking scheme, we tested all configurations for $M = 3$ and $N = 3$ relevant to specific values of $B \in 1 : MN$ by generating unique sets of configurations, $Q_{B=1}$ through $Q_{B=MN}$. We apply each set of Q_B configurations to 50 terrain examples for a specific spatial frequency, λ_{Avg}. From these example

tests, we can evaluate a configuration's expected error ($E[error_{q_h}]$) as a function of its average confidence (η_{Avg}) for a specific terrain type, λ_{Avg} (Figure 16). The tests

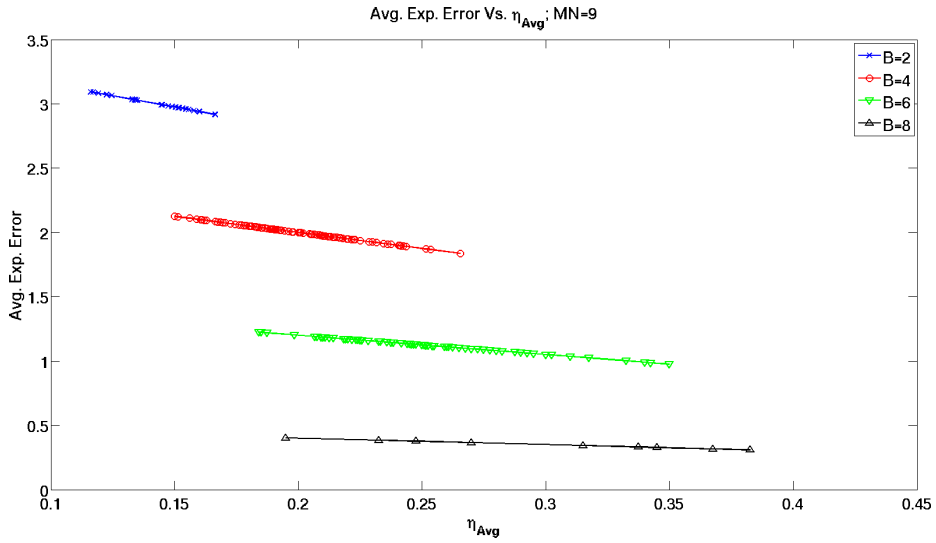


Figure 16: Expected Error ($E[error_{q_h}]$) versus Configuration Estimate Confidence (η_{Avg}), $MN = 9$, $\lambda_{Avg} = 25\%$, B=[2,4,6,8].

are repeated for $M = 4$ and $N = 4$ (Figure 17).

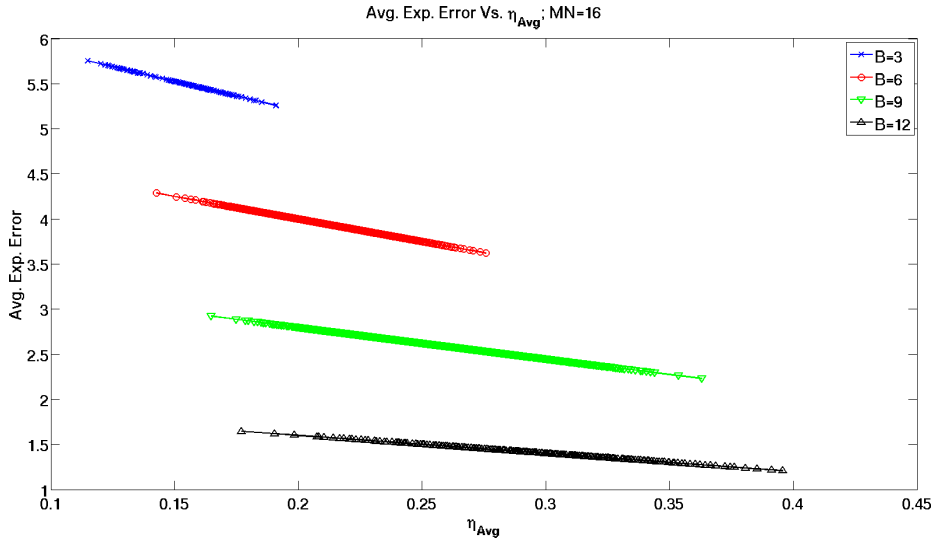


Figure 17: Expected Error ($E[error_{q_h}]$) versus Configuration Estimate Confidence (η_{Avg}), $MN = 16$, $\lambda_{Avg} = 25\%$, B=[3,6,9,12].

In the plots, each trend line represents the range of performance (average expected error) for all Q_B configurations comprised of B samples. We account for the trend

based on the linearly proportional relationship between the two quantities, η_{Avg} and $E[error_{q_h}]$ (Equations (11)-(13)).

When we isolate a single trend for a specific value of B, we can compare expected error ($E[error_{q_h}]$) versus actual error ($error_{q_h}^{Actual}$) that a set of configurations will generate when applied to terrain examples that exhibit a specific spatial frequency, λ_{Avg}. We showed, previously, that λ_{Avg} places a performance limitation on our interpolation methodology (Section 3.2.2). Setting λ_{Avg} constant for each terrain tested, we demonstrate the effectiveness of our interpolation method for a single set of Q_B configurations of increasing estimate confidence (Figures 18). For $M = 3$, $N = 3$, and $B = 4$, we plot the average expected error of all Q_B configurations and their average actual error generated when applied to 50 example terrain spaces. This data is plotted as a function of η_{Avg}.

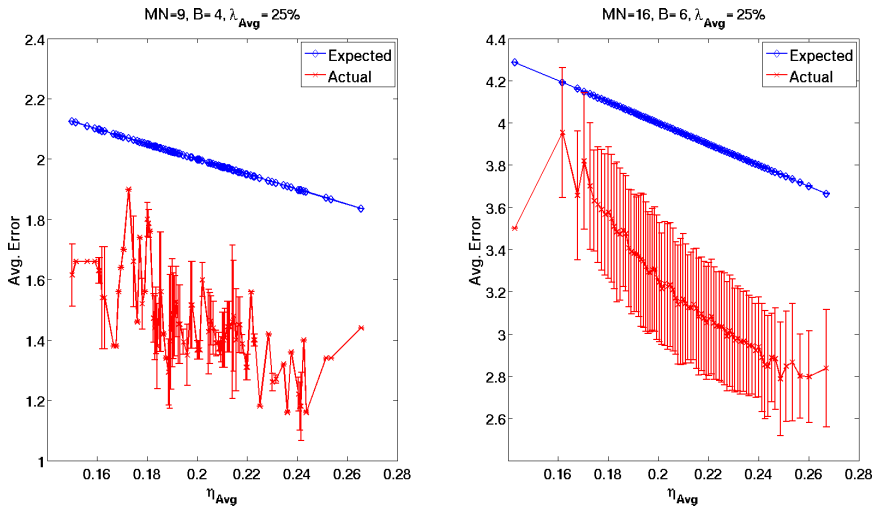


Figure 18: Error (Expected and Actual) versus Configuration Estimate Confidence (η_{Avg}), for $MN = 9$, $B = 4$ and $MN = 16$, $B = 6$, $\lambda_{Avg} = 25\%$.

In Figure 18, because of the types of terrains to which the Q_B configurations were applied, the total actual error generated falls at or below the total expected error. This trend of lower actual error is consistent across all sets of configurations for differing values of B when $\lambda_{Avg} \leq 25\%$. These tests are extended to the case of $M = 4$, $N = 4$, and $B = 6$, where we notice the same consistency in the performance

of our methodology for increasing η_{Avg}.

Using our estimation method, we can expect the upper limit on actual error, generated using these configurations, to be defined by the expected error of those same configurations. This upper limit, however, fails to hold at greater values of spatial frequency, λ_{Avg}. When the spatial frequency of the terrains increases above 50 percent, agreement between the expected error of our method and the actual error generated begins to degrade (Figures 19-20).

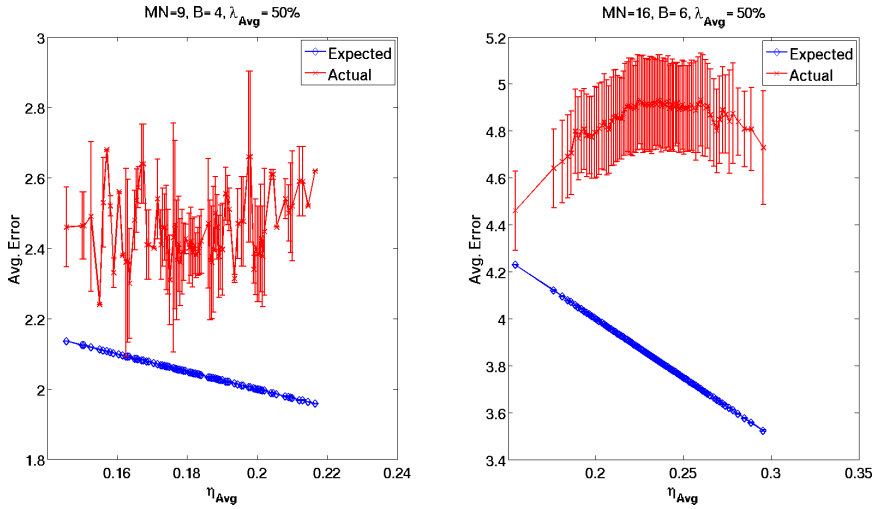


Figure 19: Error (Expected and Actual) versus Configuration Estimate Confidence (η_{Avg}), for $MN = 9$, $B = 4$ and $MN = 16$, $B = 6$, $\lambda_{Avg} = 50\%$.

Plotting expected error as a function of configuration estimate confidence enables us to hone in on the subset of Q_B configurations most capable of generating error that falls within predefined error constraints set by a scientist a priori. Knowledge of these specific configurations is useful to scientists, but also to roboticists who design navigation strategies for implementation. The final assessment needed to develop these strategies is knowledge of the cost associated with implementing each configuration, which we discuss in the next section.

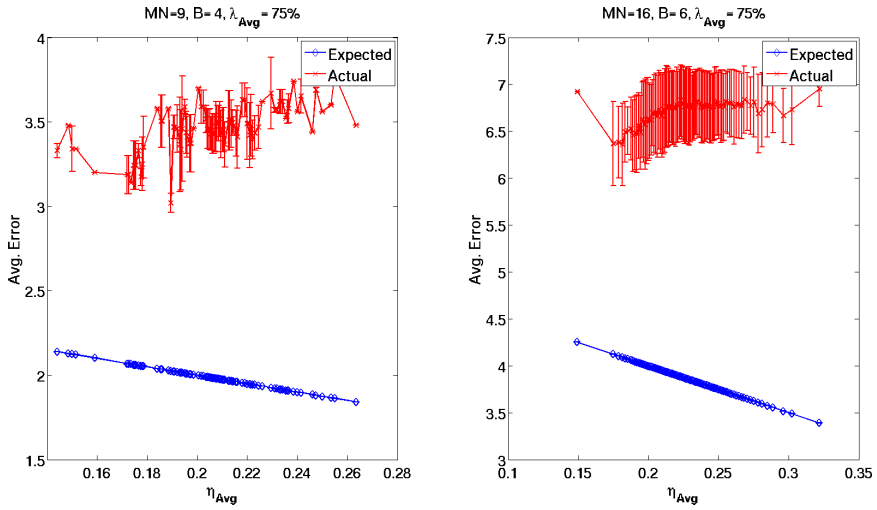


Figure 20: Error (Expected and Actual) versus Configuration Estimate Confidence (η_{Avg}), for $MN = 9$, $B = 4$ and $MN = 16$, $B = 6$, $\lambda_{Avg} = 75\%$.

3.4 Application of Performance and Resource Constraints

Thus far, we have described a methodology for informing a scientist how to quantify the range of possible sampling options available for estimating a spatially distributed phenomena. If the spatial frequency of the phenomena fall under a specified maximum limit, we have shown the performance improvement of our methodology over that of single-trial Bernoulli estimation for a configuration of any number of samples. Furthermore, we established a ranking system used to determine the estimate confidence of a particular subset of configurations relative to one another. This ranking system allows us to identify how the expected error for Q_B configurations defines a maximum ceiling on actual error, also for terrains of a specific spatial frequency. The final attribute of our work discusses the analysis of how physical constraints impact the selection of which configurations are best suited for selection and application in a space, S.

3.4.1 Average Nearest Neighbor Distance

Nearest neighbor distance information is beneficial when estimating spatial data at unknown locations (Section 3.2), but also when quantifying the spatial relationships

of known locations. Empirically-based sampling studies have shown the benefit of evaluating the performance of sampling designs as a function of nearest neighbor-based metrics. One such example is the nearest neighbor index (NNI), a ratio of the average nearest neighbor distance between all samples in a space to a baseline distance exhibited by a random spatial placement generated by a Poisson process [51]. Since our interest is in the raw cost (or sample-to-sample distance required), we leverage the average nearest neighbor distance as a measure of the resources (or cost) required to collect samples arranged in a specific configuration (Equation (20)).

$$D_{Avg}^{NN} = \frac{\sum_{b=1}^{B} d_b^{NN}}{B} \tag{20}$$

In Equation (20), D_{Avg}^{NN} is the average nearest neighbor distance calculated for a specific configuration. This metric is applied to a given set of configurations, Q_B, for a specified number of samples, B. Likewise, d^{NN} is the nearest neighbor Euclidean distance for each of the B samples in that configuration. If we plot the expected error generated by all our configurations as a function of this average nearest neighbor distance (keeping B constant to select a particular subset of configurations, Q_B, to analyze), we may identify which configurations fall within scientist-defined constraints (Figure 21). We re-illustrate our configuration constraints with the case of a space, S, defined by $M = 3$ and $N = 3$, for configurations belonging to the set Q_B, where B equals four, for terrains with $\lambda_{Avg} = 25\%$ (Figure 22). We repeat for larger spaces ($M = 4$ and $N = 4$) and B equal to six (Figure 23). As the size of the space considered grows, the density of configurations per average nearest neighbor distance measurement increases significantly such that there exists only fractions of difference in actual error between different configurations in Q_B. This increase in configuration quantity is attributed to the combinatorial relationship discussed in previous sections (Equation (6)). We plot these specific scenarios, *i.e.*, $MN = [9, 16]$, for two reasons. First, we want to represent, graphically, the relationship between the expected

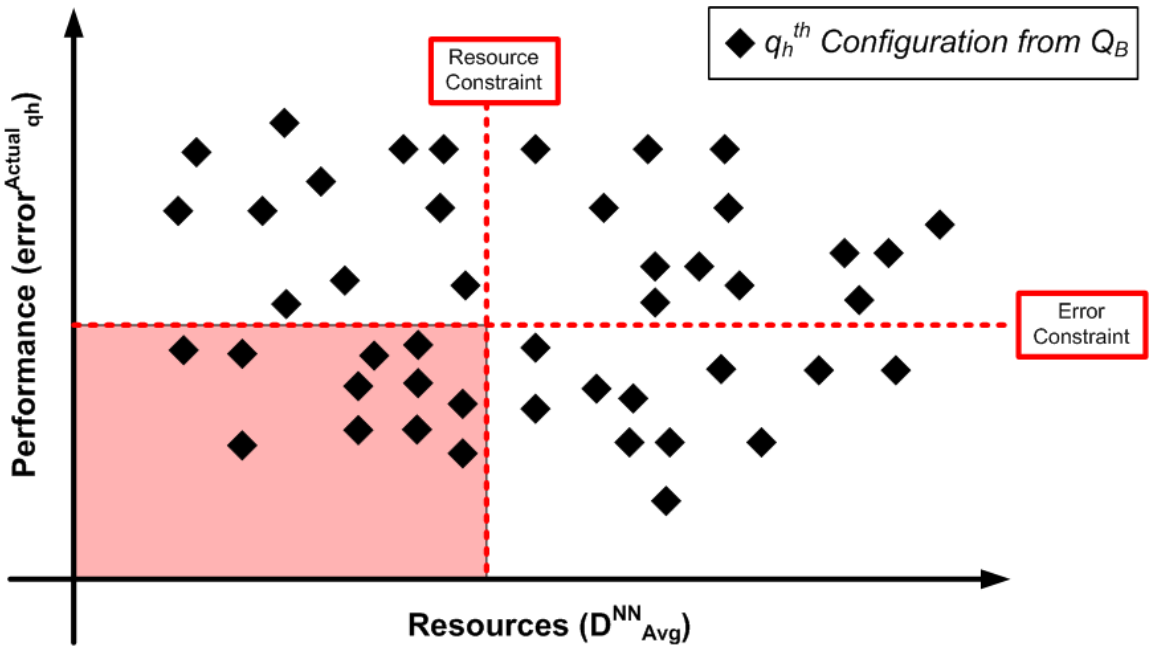


Figure 21: Illustration of how configurations are identified according to the performance constraint of error and the physical constraint of resources.

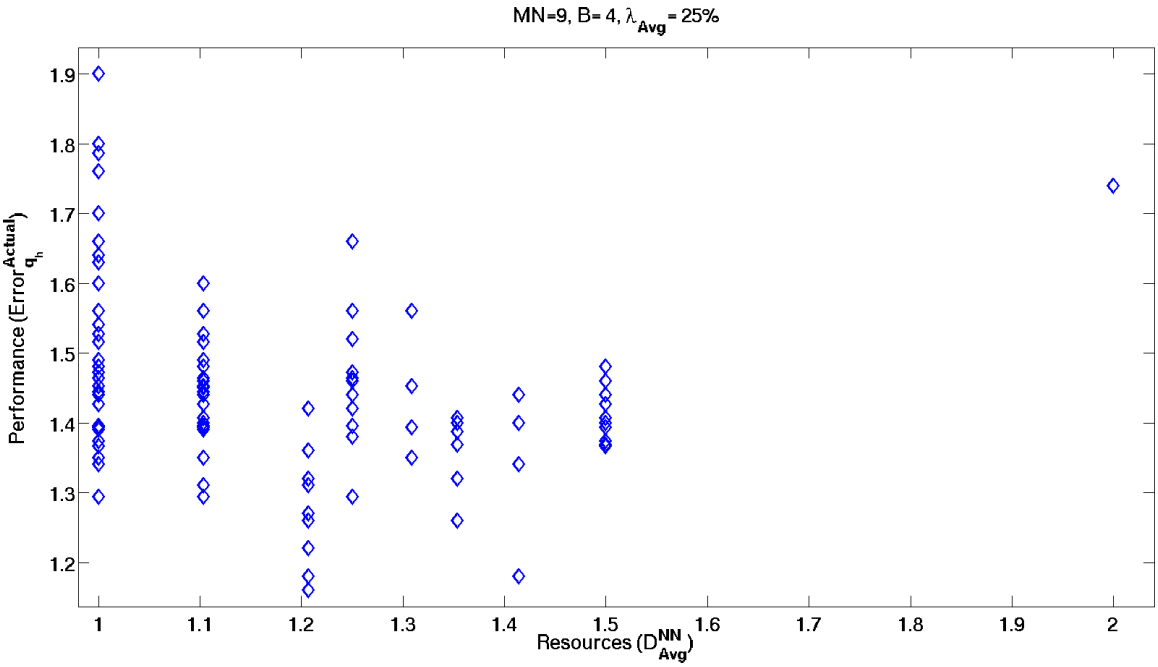


Figure 22: Actual Error ($error^{Actual}_{q_h}$) versus Average Nearest Neighbor Distance (D^{NN}_{Avg}), $(MN) = 9$, $\lambda_{Avg} = 25\%$, B=4, for h=[1..Q_B].

performance of a single sampling configuration and the resources required for that configuration to be executed. Second, we want to acknowledge the similarities that

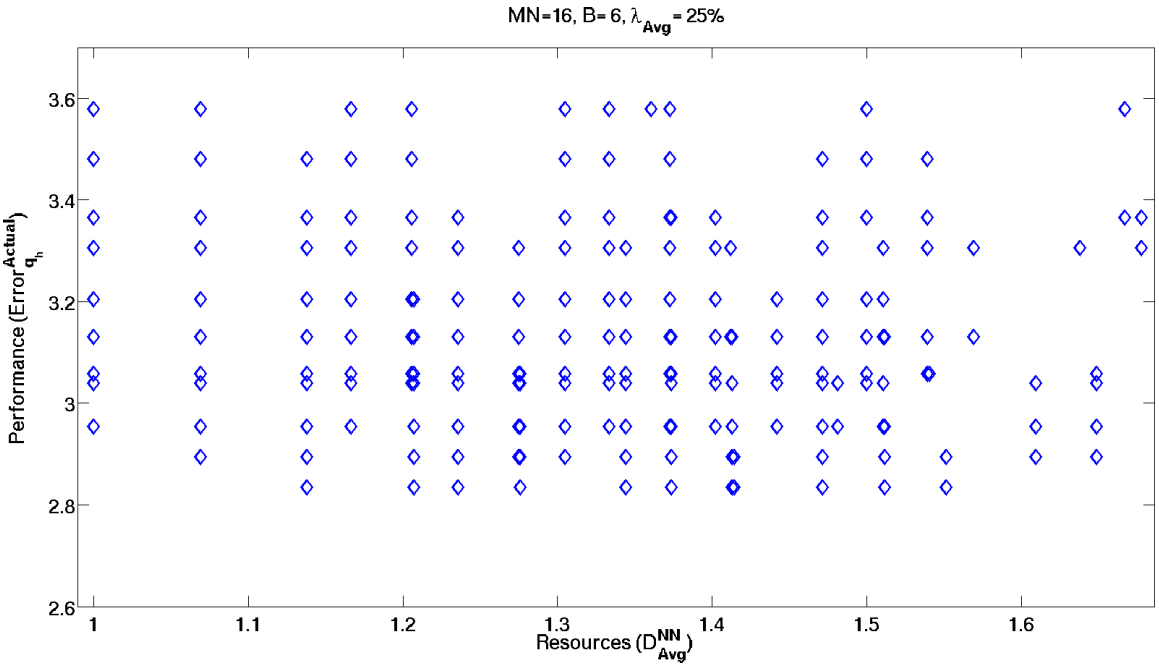


Figure 23: Actual Error ($error_{q_h}^{Actual}$) versus Average Nearest Neighbor Distance (D_{Avg}^{NN}), $(MN) = 16$, $\lambda_{Avg} = 25\%$, B=6, for h=[1..Q_B].

exist in performance and required resources between groups of configurations. These similarities are most prevalent for larger spaces where, consequently, larger quantities of configurations and larger sample sets exist. As a result, the spatial location of one or a few samples, out of B total samples, can have very little impact on the total expected error for a configuration. Yet, those same few samples can considerably impact the resources required to achieve the configuration. These two reasons are better quantified when a user defines their specific needs.

As a final comparison of which configurations (from all Q_B possible configurations) satisfy user constraints, let $error_{Max}$ and D_{Max}^{NN} represent user-defined performance and resource constraints, respectively, where $error_{Max}$ and D_{Max}^{NN} are each greater than zero. Also, let α_P and α_R be defined as the weights of user importance that are placed on each constraint being satisfied, where $0 \leq \alpha_P \leq 1$, $0 \leq \alpha_R \leq 1$ and $\alpha_P + \alpha_R = 1$. For example, if a user considers equally important that the performance and resource constraints be satisfied by a selected sample configuration, then we let

$\alpha_P = \alpha_R = 0.5$. If, however, the condition exists that the sampling configuration meeting the error constraint is more important than meeting the resource constraint, then we might let $\alpha_P = 0.8$ and α_R =0.2, yet, the specific values for α_P and α_R are user defined in addition to the constraints, $error_{Max}$ and D^{NN}_{Max}. Using these definitions, a score, ψ is generated for each configuration that falls within the boundary area defined by the contraints (Figure 21). The formula for this score is found in Equation (21).

$$\psi = \alpha_P[\frac{|error_{Max} - error^{Actual}_{q_h}|}{error_{Max}}] + \alpha_R[\frac{|D^{NN}_{Max} - D^{NN}_{Avg}|}{D^{NN}_{Max}}] \tag{21}$$

Combined with the metrics outlined in previous sections (Sections 3.2-3.3), this approach helps to establish a viable method for identifying sets of configurations that accurately estimate a spatially diverse environment while satisfying user-defined constraints.

CHAPTER IV

ESTABLISHING A NAVIGATION METHODOLOGY

4.1 Sampling Configurations via Robot Navigation

Here, we provide a background on how mobile robotics research typically addresses the task of sampling. From this added perspective, we can select a set of candidate sampling configurations relevant to robotic navigation. In Section 2.4, we presented the scope of research conducted using mobile robots for environmental monitoring applications. In the robotics community, exploration and coverage algorithms are used for obtaining complete coverage of an area where a set of sampling points are carefully chosen based on design-specific conditions. Some of these algorithms perform well with respect to their sensor-specific solutions. These algorithms highlight the novelty of incorporating a sensor in the coverage task, *e.g.*, using vision or close-range sound detection and ranging (sonar) to project subsequent waypoints for a robotic agent to follow [7, 52]. Still, these algorithms de-emphasize the importance of obtaining accurate reconstruction of the space via strategic sensor placement/measurement within the environment while operating under specific constraints.

Other algorithms consider selecting unique measurement locations, yet require data a priori to guide the exploration process and consequently, the navigation [4, 53]. This known information is typically provided by wireless sensor nodes deployed prior to an experiment for long-term monitoring [54, 55]. Still, when areas are remote and scientists are seeking survey information within unchartered locales, the importance of local navigation based on immediately sensed information is desirable for environmental monitoring applications. The following discusses the process by which we identified several navigation strategies that obtain locally-relavant detail in sampling

while requiring no prior information about the interior of the investigated area.

4.1.1 Vetting Unique Navigation Strategies

Our search for a navigation strategy that provides sufficient performance in sampling an area begins by contrasting the spiral, lawnmower-traditional, and concentric circle patterns (Figure 24) [35, 43]. Each of these patterns allows a scientist to define the

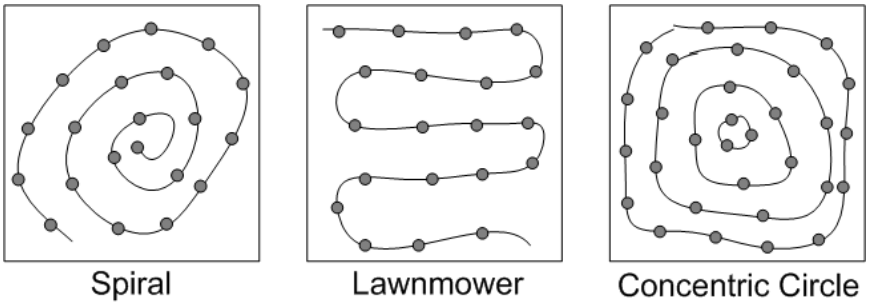
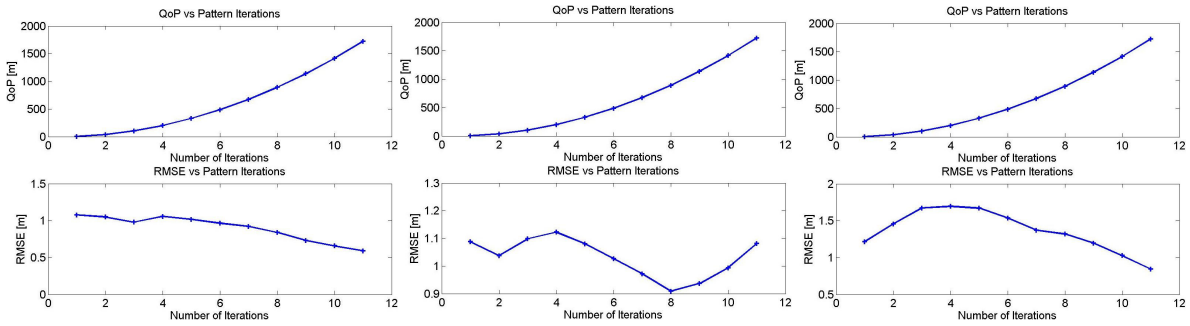


Figure 24: Illustration of first three navigation patterns considered for obtaining sampling configurations.

values of parameters (*e.g.*, swath width and iteration quantity) that will determine the range of coverage, as these parameters dictate the expansiveness of their broader geometric shape. The relevant literature contrasts the performance of these robot navigation patterns by the total Euclidean (or sequential) distance traveled by an autonomous agent executing each pattern. The metric used is the Quality of Performance (QoP), defined as $QoP = A_s/D_s$, where A_s is the total area to be surveyed and D_s is the total distance required to complete a particular coverage pattern [35]. Unfortunately, using Euclidean distance, only, can yield different root mean squared (RMS) error estimates for that same sample space (Figure 25). This is particularly true when those coverage patterns are used to designate sampling locations and resulting samples are used to reconstruct the area of interest.

As the plots demonstrate, while the impact of varying the pattern placement within the sample space is captured by the actual measure of error calculated, the

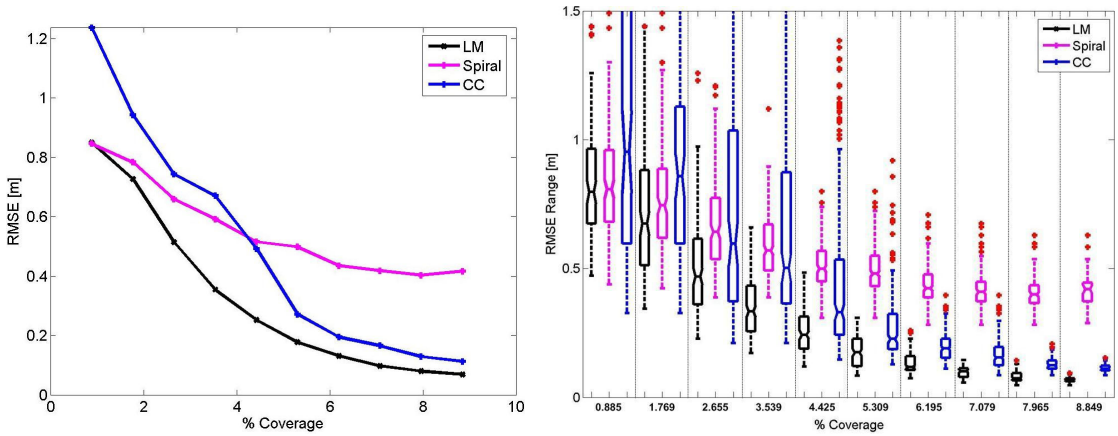


(a) Sample pattern, centered. (b) Sample pattern shifted south. (c) Sample pattern shifted west.

Figure 25: QoP (top) and RMS error (bottom) plots resulting from the simulation of an agent navigating in a *spiral* pattern. As shown, while the characteristic of measured error, *i.e.*, the RMS error curve based on collected samples, varies as a function of pattern placement within the sample space, S, the QoP does not.

variation is not reflected in the QoP. Using the QoP as the only metric in evaluating performance of a robotic survey system presumes the sample selection process is independent of the phenomena being measured. Also, when adapting navigation to changes in monitored phenomena, symmetric application of a navigation pattern can not be guaranteed, thus highlighting the value in using metrics that are specifically designed to account for variable information (Chapter 3). Refer to Tunstel *et al.* and Parker *et al.* for a discussion of the theoretical [35] and practical [14, 56] implementation of lawnmower-based navigation patterns considered for robotic surveying. In light of this ambiguity in extracted information about the surveyed area, we focus on the error metric as our primary tool for contrasting performance of these navigation patterns as sampling schemes. To compare these patterns, we measure the RMS error, $\hat{\theta}$, between the original terrain and terrain re-created from samples collected in the shape of each pattern. Also, we reconstruct these spaces using interpolation options designed for estimating continuous spaces. The RMS error metric is used to measure accuracy of estimates, especially in the mapping profession, and is relevant for our initial comparisons of sampling schemes. We will discuss more about the usefulness of RMS error and our interpolation selection for continuously-valued data in the next chapter.

To ensure sufficient coverage, the robotic agent executes each pattern symmetrically across the sample space, with successive swaths evenly placed. For evaluation, we execute the three patterns (lawnmower-traditional, spiral, and concentric circle) on a set of 100 digital elevation models (DEM) with different undulating features that mimic realistic terrains (`www.gameprogrammer.com/fractal.html`). By evaluating each pattern at different percentages of coverage, we calculate the average RMS error along with the variability in that error (Figure 26).



(a) RMS error.

(b) Variability in RMS error.

Figure 26: Contrasting RMS error between lawnmower-traditional, spiral, and concentric circle navigation patterns.

Consistently across all ranges of percent coverage, the lawnmower-traditional pattern yielded the lowest average RMS error. The traditional definition of percent coverage, *i.e.*, total Euclidean distance divided by total distance between all possible sampling locations, was used to sort and plot the results in Figure 26. As coverage increases, the error generated by lawnmower-traditional and concentric circle patterns begins to converge, yet, this is commonly the case for any pattern when sampling quantity, and thus coverage, increases. Since our interest is in identifying the navigation strategy that minimizes RMS error, we remove the spiral and concentric circle patterns from further consideration, focusing on developing navigation strategies that will lower the average RMS error beyond the performance of the

lawnmower-traditional pattern (Algorithm 1).

Algorithm 1 Lawnmower-traditional navigation scheme

Require: Static navigation policy, $r(q)$ ($q =$ *All northing positions across a swath*)
Require: Border dimensions, $dimx$, $dimy$; Sample space, S
$(P^{SL}) = S/Resources$ {Define vector of starting locations for each swath according to resources available}
for $k \leq length(P^{SL})$ **do**
$X^{Total}(P^{SL}(k), all) = S(P^{SL}(k), all)$ {Navigate across S, collecting all samples at designated Y-position, $P^{SL}(k)$}
end for{Store total set of samples}
return X^{Total}

Initially, we proposed that by applying methods influenced by informed search techniques, we would succeed in obtaining even lower RMS error results than the lawnmower-traditional pattern. This first attempt at "intelligent" navigation was formulated as guiding the agent's navigation decisions according to ad-hoc combinations of steepest ascent and steepest descent rules. For simulation testing, we assumed the availability of multiple agents, where each agent is assigned an initial location on the border of the sample space, S, and proceeds to navigate according to the aforementioned control laws. As a part of the navigation design, these starting points were defined only at evenly distributed locations around the border of the sample space as we do not presume knowledge of the area's interior. Since part of our contribution is predicated on operating with a lack of a priori knowledge, the border was used as our only information reference. Furthermore, the evenly distributed border locations were a better option since allowing each agent to start at a location in the same region led to poor performance. This was evident when we discovered that each agent would get caught in the same local minimum or maximum, failing to cover areas remotely located from their start position. We observed, in executing these navigation policies, that applying solely one rule (*e.g.*, steepest ascent or steepest descent) limited the extent of internal penetration and sample collection of the area of interest. We, therefore, tried alternating combinations of both of these policies. Initially, we

instructed agents to navigate according to a steepest ascent policy. Once a peak in local maxima was reached for all agents, a new border surrounding the remaining unsampled, inner portions of the sample space was defined and the policy was repeated with new "psuedo-border" locations allowing increased interior penetration of the sampling area. Navigation was resumed using the steepest descent policy to achieve increased internal penetration of the space (Algorithm 2).

Algorithm 2 Combined navigation policy procedure.

Require: Choice of primary and secondary navigation policies, $r_1(q)$ $\&r_2(q)$ (q = *Steepest Ascent* || *Steepest Descent*)

Require: Number of agents, N ($N \geq 1$); border dimensions, $dimx$, $dimy$; Sample space, S

{Define the required number of starting locations on border for N agents.}

$border = locate(S)$ {Store a vector of all the (x, y) coordinates for each possible border location in terrain S.}

for $n = 1$ to N **do**

$S_n(x, y) = border(n)$ {Define the next agent's starting location indexed according to n.}

$X_s = Apply\ r_1(q)\ at\ S_n(x, y)$ {Apply primary navigation policy to terrain and store recorded samples.}

end for

Require: border dimensions of the interior, $dimx_i$, $dimy_i$; Unsampled interior of S, S_i

{Define the required number of starting locations equal to the number of border locations}

$border_i = locate(S_i)$ {Store a vector of all the (x, y) coordinates for each possible border location in terrain, S_i.}

for $n = 1$ to $length(border_i)$ **do**

$S_n(x, y) = border_i(n)$ {Define the next agent's starting location indexed according to n.}

$X_s^i = Apply\ r_2(q)\ at\ S_n(x, y)$ {Apply secondary navigation policy to terrain and store recorded samples.}

end for

$X^{Total} = X_s + X_s^i$

return X^{Total}

An example output of the sampled space is shown in Figure 27.

In similar fashion to the comparisons between lawnmower-traditional, spiral, and concentric circle patterns, we placed the lawnmower-traditional pattern alongside our next set of potential navigation methods, termed SA-SA (steepest ascent-steepest

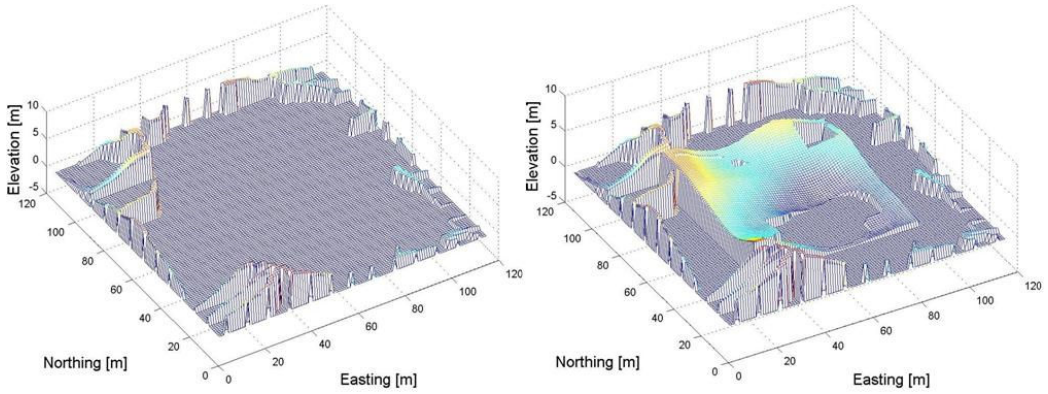
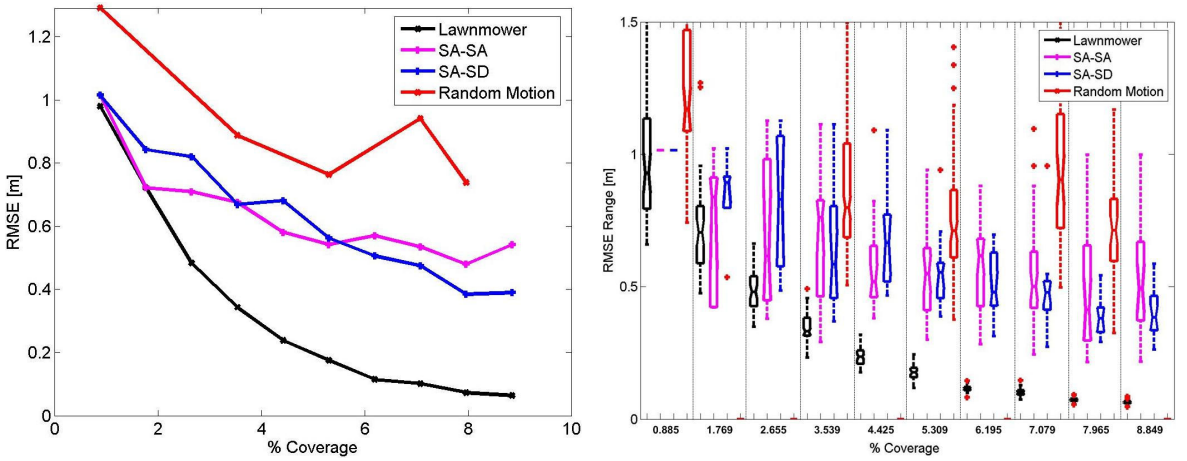


Figure 27: Two-phase application of simple control laws on sample space, initial application of steepest ascent rule(left), subsequent application of steepest descent rule (right).

ascent), SA-SD (steepest ascent-steepest descent), and random motion. The latter method, random motion, was included to test if providing arbitrary, or random, heading information would yield sample configurations with lower RMS error than both a static approach, *i.e.*, lawnmower-traditional, and a more informed one, as found in our SA-SA and SA-SD navigation policies. Surprisingly, each method fell short of outperforming lawnmower-traditional, based on the RMS error plots generated (Figure 28).



(a) Average RMS error.

(b) Variability in average RMS error across 50 different simulated terrains.

Figure 28: Contrasting RMS error between lawnmower, SA-SA, SA-SD, and random motion navigation patterns.

There were multiple problems with the SA-SA and SA-SD approaches. First, blindly applying a rule-set to sensed data is too reactive of a behavior for a system where minimizing resource usage is a design goal. Part of our goal in developing an adaptive navigation scheme is having the capability to control the robotic surveying system (RSS) to ensure performance objectives can be met even when these resources are minimized. If specific constraints are not placed on the agent's execution of these control laws, the upper bounds on the resources are no longer deterministic and become purely driven by the phenomena. Since informed search methods influenced the design of these navigation strategies, our heuristics provided agents with a series of sub-goals, locating local minima and maxima, without an upper bound on how many of each should be achieved. Second, there is no spatial structure in the samples collected as the direction of the paths generated are arbitrarily dictated by the agent's adherence to its rules of navigation. This lack of control over sample placement could leave certain areas more heavily sampled than others. Take, for example, the heavily sampled northwest interior corner of the test DEM versus the void of samples in the southeast corner in Figure 27. Although four possible combinations of the steepest ascent/descent policy were tested (*i.e.*, SA-SA, SA-SD, SD-SA, and SD-SD), only two are presented in this preliminary analysis as an example since none of the four options led to lower RMS error than lawnmower-traditional. This test was repeated across the same test set of 100 DEMs and the RMS error values were averaged, showing no improvement over the lawnmower-traditional navigation strategy.

4.2 Establishing a Baseline Navigation Method

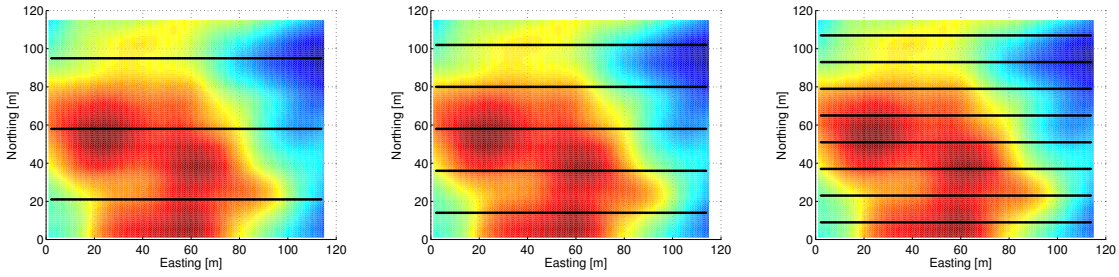
As a result of the preliminary navigation testing, we turn our attention to the lawnmower-traditional navigation pattern and the variations available that produce a viable set of sampling configurations. This portion of our work is inspired by the

hydrographic survey experiments discussed in Section 2.4 [7]. We chose a set of navigation solutions that each yield specific performance and resource constraint-based benefits. Additionally, these solutions leverage the benefit of collecting information symmetrically (*i.e.*, executing even, parallel swath-like trajectories across the area). For each solution in the set, how the characteristics of their trajectories change is defined based on the information collected online during navigation.

4.2.1 Lawnmower-traditional navigation

Research in the field of environmental monitoring can place heavy emphasis on the necessity of only considering "intelligent solutions", however, previous work has shown the sufficiency of drafting and executing a "static" navigation plan without making concessions for adaptation during execution [7, 42]. The primary static navigation scheme selected is lawnmower-traditional navigation (Algorithm 1).

Just as it sounds, lawnmower-traditional navigation assumes the execution of repetative, evenly-spaced, linear trajectories across an area. For our purposes, we dictate that each sample, collected by an agent executing this navigation scheme, is positioned along a straight line as a part of a particular swath (Figure 29). The



(a) Max swath width: 14 meters (b) Max swath width: 22 meters (c) Max swath width: 37 meters

Figure 29: Lawnmower-traditional trajectories that provide static, linear data collection, preserving resource usage but achieving minimal spatial diversity of samples. Navigation is in the positive x direction for successive swaths placed along the y axis (3, 5, and 8 swaths).

primary variables of interest dictating the lateral diversity of the entire pattern are the

size of the space, M, N, and the user-defined resources. These resources allow a user to project how many linear swaths, totaling B samples, may be executed throughout the space, S. For example, if a scientist only has enough battery power, *i.e.*, resources, to sustain navigation across the sample space two times, then two reference swaths evenly spaced across the area will be applied and two swaths of samples will be collected. The lawnmower-traditional navigation is a conservative sampling scheme, minimizing resource usage, *i.e.*, an agent traverses short distances between subsequent samples. In contrast, the performance constraint (*i.e.*, reconstruction error) is more dependent on the data observed at each sampling location and the estimation methodology selected.

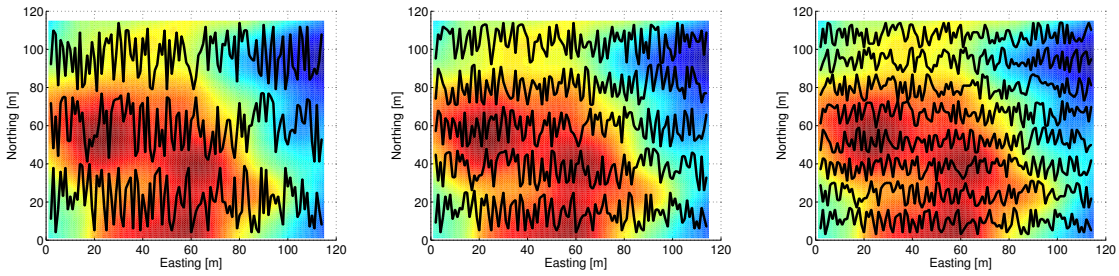
4.2.2 Lawnmower-random navigation

We include a second sampling methodology to test the feasibility of previous approaches in randomness considered by both the sampling and robotics communities [1, 5, 18, 20, 53]. The concept of random sampling for our application is implemented by defining the agent's heading towards its next sample along a swath according to a randomly selected value in situ (Algorithm 3).

Algorithm 3 Lawnmower-random navigation scheme

Require: Random navigation policy, $r(q)$ ($q =$ *Random variable*)
Require: Border dimensions, $dimx$, $dimy$; Sample space, S
$(P^{SL}, bw) = S/Resources$ {Define vector of starting locations for each swath according to resources available.}
{Define bandwidth (*i.e.*, swath width) of allowable search range according to resources available for the sample space.}
for $k \leq length(P^{SL})$ **do**
$Y_{Range} = [P_k^{SL} \pm bw]$
$X_s = X_s + Apply\ r(q)\ at\ S_{(x,y)}^{P_k^{SL}}\ for\ (1 \leq x \leq dimx, y = Y_{Range})$
{Apply random navigation policy to terrain along a single swath and store recorded samples.}
end for
$X^{Total} = X_s$ {Store total set of samples}
return X^{Total}

This method provides a spatially-diverse sampling distribution and would appear to be an ideal navigation strategy with respect to the performance constraint, *i.e.*, low reconstruction error. The one caveat to this approach is that the robotic survey system must actually navigate to these randomly selected waypoints and, therefore, will likely incur a hefty penalty in total distance traveled. This increase in distance traversed compromises the opportunity for lawnmower-random to successfully meet a user-defined resource constraint, D_{Max}^{NN}. Example lawnmower-random sampling trajectories, for an increasing swath count (or B samples), are shown (Figure 30).



(a) Max swath width: 37 meters (b) Max swath width: 22 meters (c) Max swath width: 14 meters

Figure 30: Lawnmower-random navigation trajectories that guarantee well-distributed sample location, yet are costly in terms of resources. Navigation is in the positive x direction for successive swaths placed along the y axis (3, 5, and 8 swaths).

4.2.3 Lawnmower-informed navigation

Like lawnmower-random navigation, making spatially-dynamic sampling decisions online also has promising benefits to improving performance, *i.e.*, lowering reconstruction error. Simultaneously, in contrast to the static nature of lawnmower-traditional, there are benefits to accomplishing the sampling task without expending unnecessary resources. When specific information is detected in sampled information, *i.e.*, better information at lower cost, a greater reward relative to a global goal should cause a change to navigation decisions. Previous empirical tests we performed had shown that one feature common to maps that yielded error values lower than those sampled

according to the lawnmower-traditional pattern was the presence of partial, continuous paths scattered throughout the terrain. This is most likely the case because of the variety in sample placement in contrast to the stark linearity and uniformity of the samples collected according to Algorithm 1. We recommend a navigation strategy whose waypoints are both non-deterministic and do not require traversing long distances. With this navigation approach, we obtain greater spatial diversity of sampled information like lawnmower-random, yet with the specific aim of using our informed estimation theory (Chapter 3) to influence subsequent sample selection. Additionally, sample selection is influenced by proximity of potential samples to the robot's current position, to attempt at reducing resource usage while improving reconstruction error (Algorithm 4).

Algorithm 4 Estimation-informed navigation scheme

Require: Choice of navigation policy, $r_s(q)$ ($q = \eta_{Avg} = estimation\ confidence$)
Require: Border dimensions, $dimx$, $dimy$; Sample space, S
$(P^{SL}, bw) = S/Resources$ {Define vector of starting locations for each swath according to resources available and bandwidth (*i.e.*, swath width) of allowable search range according to resources available for the sample space.}
for $k \leq length(P^{SL})$ **do**
$Y_{Range} = [P_k^{SL} \pm bw]$
$X_s = X_s + Apply\ r_s(q)\ at\ S_{(x,y)}^{P_k^{SL}}\ for\ (1 \leq x \leq dimx, y = Y_{Range})$
{Select subsequent sample based on 1) the estimation confidence, η_{Avg}, and 2) proximity to agent's current position calculated for each location in a candidate set of samples.}
5: **end for**
$X^{Total} = X_s$ {Store total set of samples}
return X^{Total}

As with lawnmower-traditional navigation, the actual number of swath references used is a function of the user-defined resources. We presume our agent's general heading to be directed in the positive x direction, using static lawnmower swath locations as a reference. Meanwhile, it is intended for the agent's actual heading to deviate from this reference as a function of two values calculated for each sample from a candidate set at each time step. These two values are the estimation confidence,

η_{Avg}, and the distance from the agent. Each sample selection is made, *i.e.*, the function, $r_s(q)$ is applied, specifically, by identifying which location, from a local subset, generates the poorest confidence value, η_{Avg}, and is physically located the closest to the robot's current location. We identify candidate sample locations generating the poorest confidence value so that we may reduce the quantity of unvisited locations whose expected error will be greatest. At the same time, we filter from those locations with the poorest confidence the sample location closest to the robot's current location to reduce resource usage. This step in our algorithm is analogous to a conservative exploration policy, aiming to achieve minimal error at a reduced cost. The confidence, η_{Avg}, is calculated at each time step based on the current samples already included in the total set of observations to that point in the navigation. The sample location, from the candidate set, found to meet this criteria is selected as the next sample (Figure 31-34).

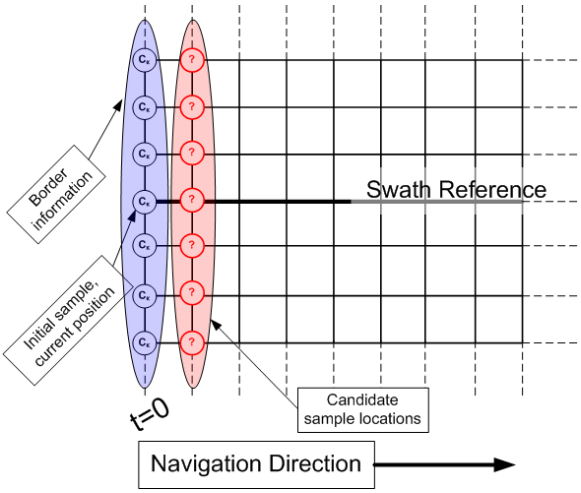


Figure 31: Initial scene for sample selection, including initial border information designated as c_κ, where κ equals 1 or 2 in the dual-class environment.

These steps are repeated until the end of the designated reference swath is reached within the sample space (Figure 35-38). When this series of sample selection decisions is executed during navigation, a specific set of trajectories is generated (Figure 39).

Our estimation methodology could be implemented multiple ways to accomplish

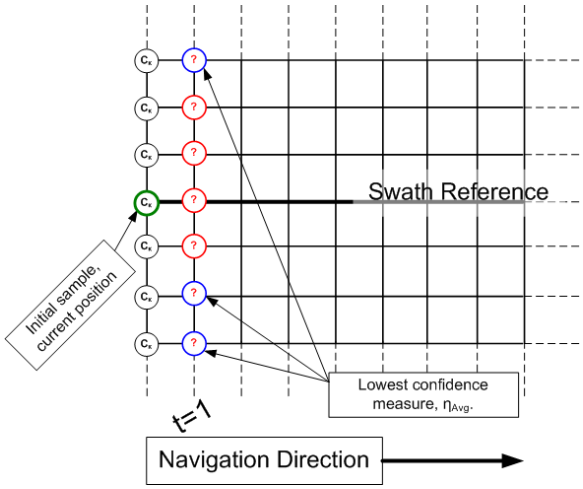


Figure 32: Identify potential sample locations with lowest confidence measure, η_{Avg}.

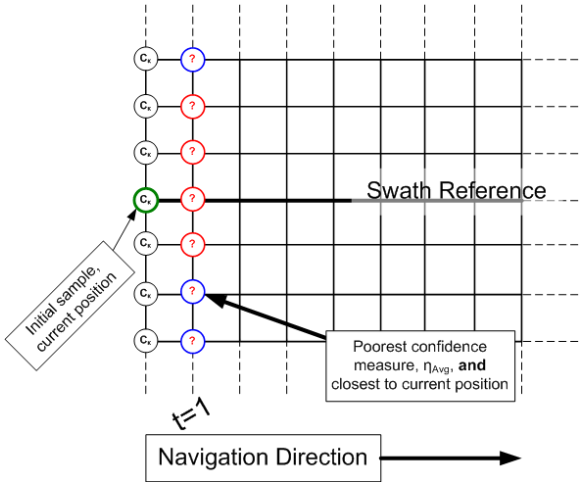


Figure 33: Identify potential sample locations with lowest confidence measure, η_{Avg}, located closest to current location.

sample selection. We present this form of a decision policy as an example of navigation that is specifically driven by our informed estimation theory.

As example navigation options, each type discussed here (lawnmower-traditional, lawnmower-random, and lawnmower-informed) yields a unique approach to achieving a spatially diverse set of sampling configurations. In the next chapter, inspired by these approaches to navigation, we validate each of these robotic-based sampling schemes with our metrics from Chapter 3. We will also identify how these sampling schemes are compared against one another according to performance and resource

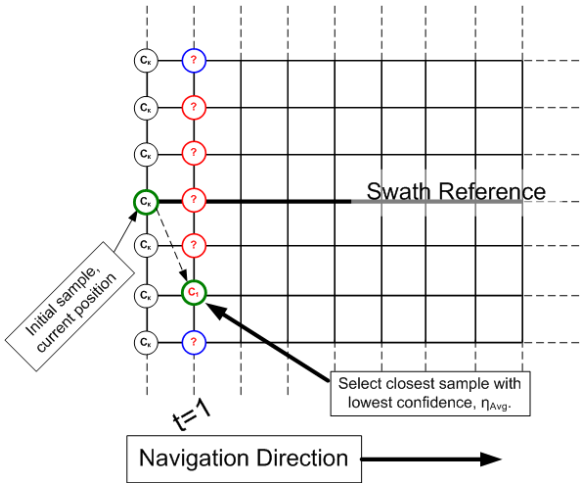


Figure 34: Make sample selection from options satisfying lawnmower-informed algorithm criteria.

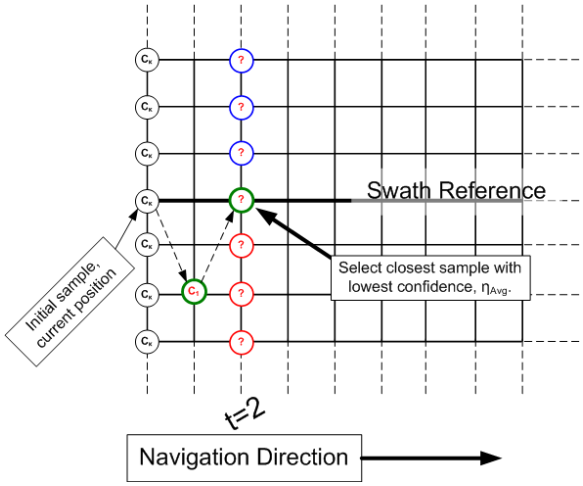


Figure 35: Repeat of sample selection at t=2.

constraints.

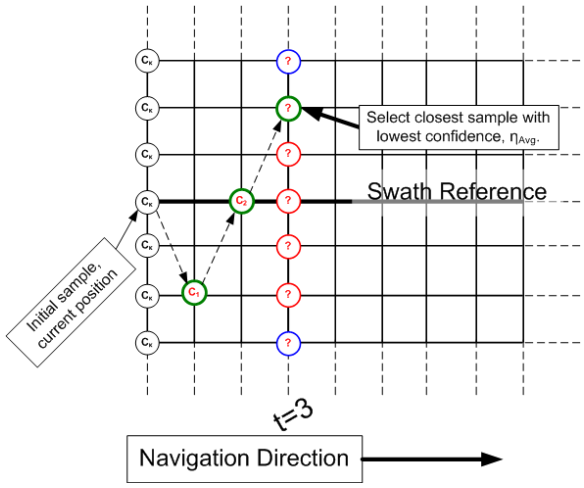


Figure 36: Repeat of sample selection at t=3.

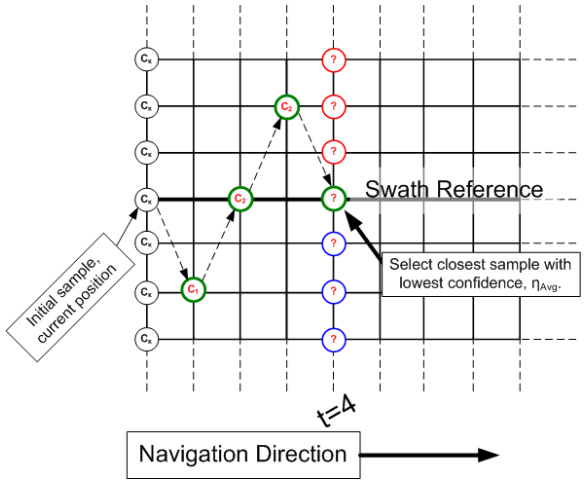


Figure 37: Repeat of sample selection at t=4.

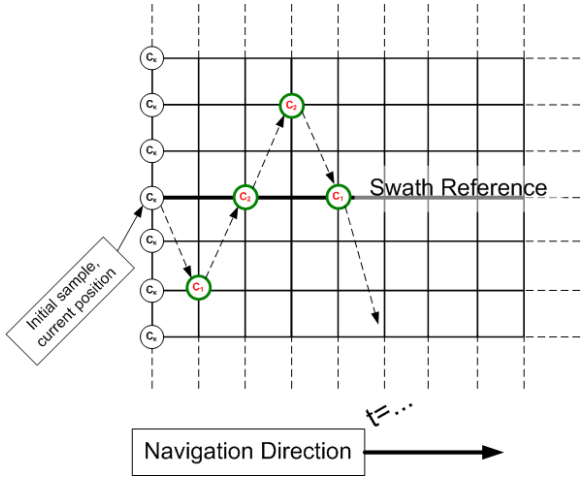
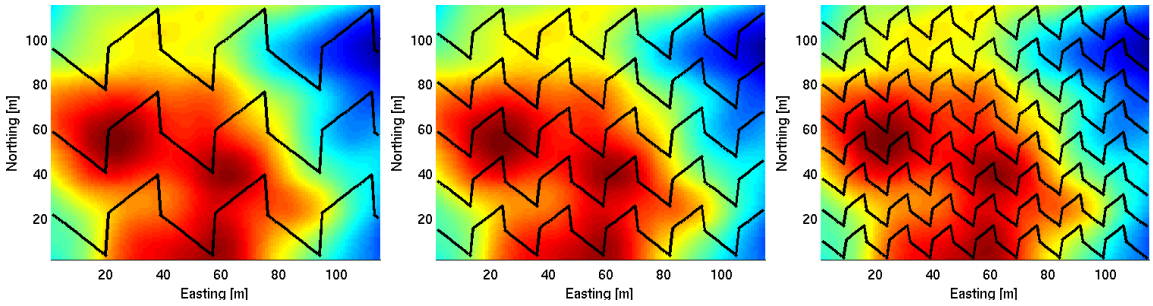


Figure 38: Repeat until end of reference swath is reached and navigation path is dynamically created.



(a) Max swath width: 14 meters (b) Max swath width: 22 meters (c) Max swath width: 37 meters

Figure 39: Informed navigation trajectories that provide increases in spatial diversity. Sampling approach preserves resource usage but also makes intelligent sampling decisions as a function of our informed estimation methodology. Navigation is in the positive x direction for successive swaths placed along the y axis (3, 5, and 8 swaths).

CHAPTER V

EXPERIMENTAL SETUP AND RESULTS

In Chapter 3, we outlined our theory for our informed estimation method and evaluating the performance of that method when applied to a space, S. In Chapter 4, we used this theory to design a navigation scheme for the purpose of sampling, also introducing other baseline navigation algorithms. Earlier, we confined the validation of our theory to a small configuration space based on the exhaustive set of all Q_{Total} sample configurations in those spaces ($MN = [9, 16]$, Section 3.2.1). In this chapter, we validate our method for robotic navigation algorithms that generate a set of sample configurations within more realistically sized spaces. The size of these spaces are more relevant to the actual sizes of the pixel footprint discussed in Chapter 2 (Section 2.1). We, first, introduce our system analysis tools, including a vetted digital elevation map generator (or "DEM maker"), used to validate these navigation algorithms as viable sample selection mechanisms.

5.1 Modeling Spatial Data: Simulation and Validation Tools

We employ several tools that assist in validating our sampling and estimation methodology. Metrics and technical specifications relevant to mapping are the first item to address.

5.1.1 RMS Error Metric in Spatial Sampling

Concerning evaluating the quality of our simulation tools and the generated terrains, we define the performance metric root mean squared (RMS) error, $\hat{\theta}$. This metric is expressed in Equation (22), where Z_0 is our truth data and $\hat{Z}_0$ is the estimated data generated as a result of the samples collected according to a selected navigation

policy.

$$\hat{\theta} = \sqrt{E[(Z_0 - \hat{Z}_0)^2]} \tag{22}$$

The one caveat to using RMS error is its sensitivity to outlier data typically found as a result of measurement noise [26]. Although this is a valid concern, our interest is soley in the error of values at unvisited locations of our simulated environments. Additionally, the spatial interpolator and filtering process selected allows us to control the presence of such outliers in our data set. Finally, in Section 2.3.1, we discuss the tradeoffs between the cost of the sampling operation, itself, and the cost of navigating to a designated sampling location; the former is a scenerio that would require an error model for each measurement made. In our DEM maker, however, our focus is on providing an accurate terrain model only. Quantifying error in the form of RMS error is a convenience, as it provides a single error estimate, which is a function of the value at each location in the sample space, but also representative of the entire area. This succinct value allows us to test a larger number of digital elevation maps (DEMs) generated, and make sound observations about the ability of our DEM maker to faithfully represent a heterogeneous spatial data set.

5.1.2 Mapping Accuracy Standards

One of the specific phenomena used to validate our simulation tools is elevation. As such, we evaluate the quality of our DEM maker, in part, by considering accepted accuracy standards employed by professionals in the cartographic and photogrammetry fields. Although these standards are applied to maps of scales much greater than our in situ mapping tool, we refer to them for perspective. The three most popular standards are the National Map Accuracy Standard (NMAS), American Society for Photogrammetry and Remote Sensing (ASPRS) standard, and the National Standard for Spatial Data Accuracy (NSSDA). NMAS is currently considered a legacy metric since the technology available to produce high-resolution maps was limited

during its inception and larger mapping errors were more acceptable. ASPRS and NSSDA are more closely adhered to given the advancements in photographic sensors and stereoploters [57]. These two standards differ exclusively in their usefulness as a quantitative reference to which mapping products can compare. ASPRS provides (in actual units), scales of accuracy, whereas NSSDA provides a way of stating accuracy in terms of the RMS error calculated. The two standards are tabulated for maps with different contour intervals (Table 1, adapted from [57, 58]).

Table 1: ASPRS and NSSDA map accuracy standards (vertical), $\zeta = 1.96$.

Horizontal Contour Interval [m]	Vertical ASPRS-RMS error [m]	NSSDA (ζ*RMS error) [m]
0.3048 (1 [ft])	0.0508	0.0996
0.6096 (2 [ft])	0.1016	0.1991
0.9144 (3 [ft])	0.1524	0.2987

The statistical measure of accuracy for NSSDA is based, in part, on a technical report prepared by Greenwalt and Schultz [59]. It outlines the application of probability theory to cartography and includes appendices for computing vertical and horizontal accuracies in terms of their RMS error. The NSSDA adopts a specific confidence interval (95 percent) which designates, according to a table of linear standard errors, a multiplier, ζ, to be applied to the RMS error calculated. Specifically, data reported according to NSSDA states that a map will not have error at any single location in excess of 1.96*(RMS error) with a 95 percent confidence level. This practice is cited multiple times in the NSSDA literature [58]. In contrast, the ASPRS standard requires that RMS error values be calculated at well defined points and any contour map of the data must be produced with intervals in accordance with the limits in Table 1. The subjectivity of what constitutes well-defined points on a map must be

resolved by all those invested in the mapping product.

While the NSSDA is useful for contractually bound projects between parties, it serves only as a reference with respect to scientific study. For our application, responsibility defaults to the user for defining acceptable accuracy limits. Being that our focus is on capturing detail throughout the sample space more so than providing a reference map for public consumption, we will reference the expectations of the ASPRS standard. For all our work, we evaluate the RMS error results presented in our preliminary data in later sections based on Class 1 vertical accuracy, *i.e.*, the accuracy reported in Table 1. Class 2 and Class 3 vertical accuracy requirements are determined by multiplying the horizontal contour interval and associated vertical ASPRS-RMS error values for Class 1 by a factor of two and three, respectively.

5.1.3 Creating Simulated Digital Elevation Maps

In simulation, we generate digital elevation maps (DEM) capable of representing realistic terrain in both quantized and continuous formats (Section 3.1). In the continuous space, these terrain examples are analogous to areas like the Arctic or places that are rarely accessible for human exploration without more advanced data collection methods such as airborne or satellite surveillance (Figure 40).

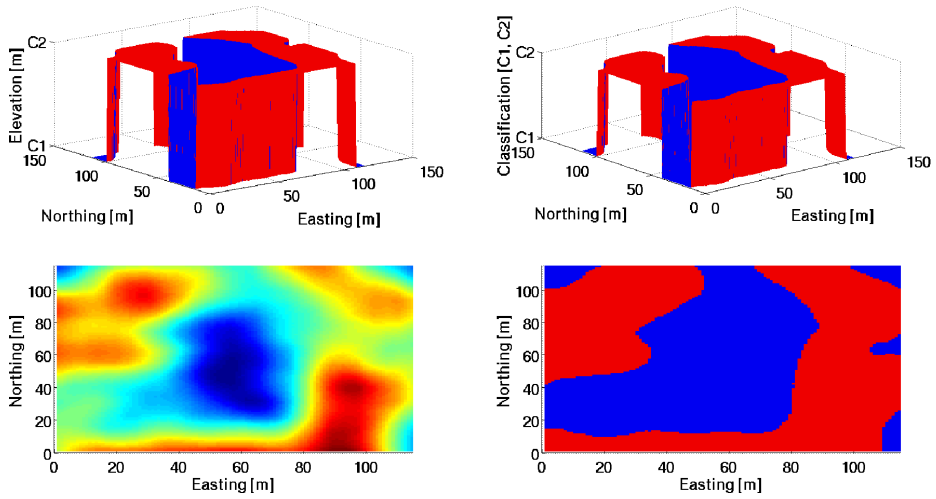


Figure 40: Simulated DEMs in both quantized and continuously valued formats.

With this DEM maker, we are able to vary the placement of smooth hills, jagged craters, and other forms of spatial anomalies within an empty space, or generate completely random, spatially complex terrains altogether. The latter is easily accomplished with popular computer graphics techniques [60]. This allows us to project the types of detail existing within the pixel footprint of remotely sensed data (Section 2.1). In addition to purely synthetic DEMs, we can translate realistic areas provided in the form of 2D contour maps into viable DEMs for simulation purposes. To accomplish this translation from 2D contour to 3D maps, we require access to reference data in a usable format so that quantitative comparisons to DEMs generated based on information collected with our robotic survey system can be made. This contour information is available in a limited capacity in the form of maps provided by local and state entities, as the 3D elevation data is proprietary. The basic steps employed by our 2D-to-3D mapping tool are outlined here and depicted visually in Figure 41.

1. Import a 2D contour image.

2. Associate the (x, y) coordinates of the contour pixels with their appropriate elevation value.

3. Interpolate between the contour values to estimate (x, y) coordinates and generate a complete map.

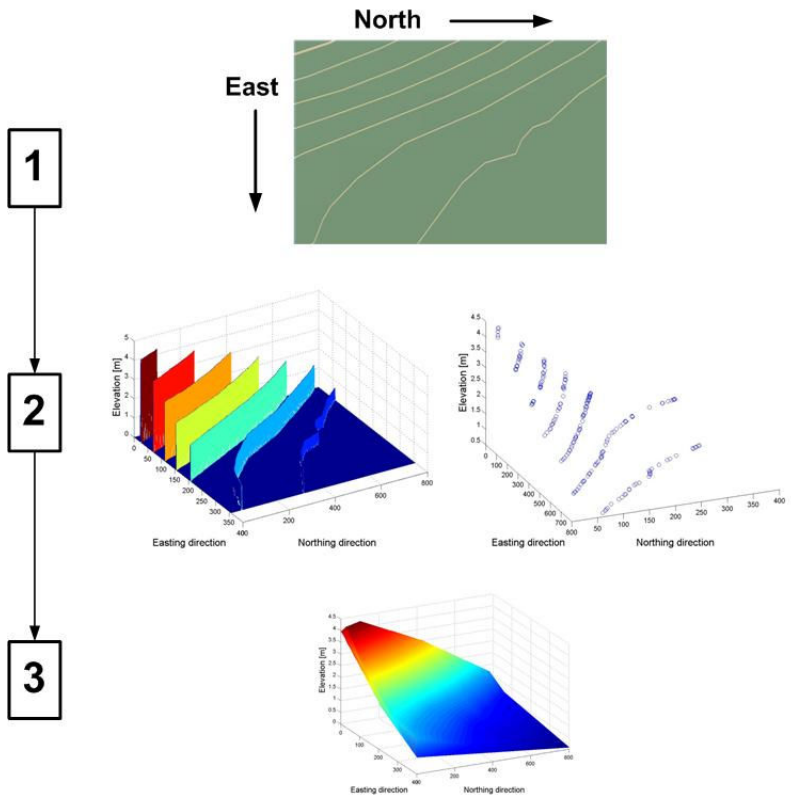


Figure 41: 2D-to-3D map generation.

The series of steps required to translate these topographic maps to usable digital elevation models includes image processing, assigning unique elevations to identified contours, and interpolation between those elevation values. Typically, image enhancement techniques are needed to compensate for the low quality of the contour maps and the variations in contour line thickness that exist (Figure 42). A 2006 lidar survey of select areas was conducted by the city of Atlanta's Department of Watershed Management and provides the publically available contour data from which we extract specific areas of interest.

5.1.3.1 Contour line extraction and noise reduction

The first step in converting 2D contour data into 3D terrain maps requires color normalization. We first convert the raw contour image of our selected area to grayscale. After considering different edge detection schemes [61–63], individual contour lines

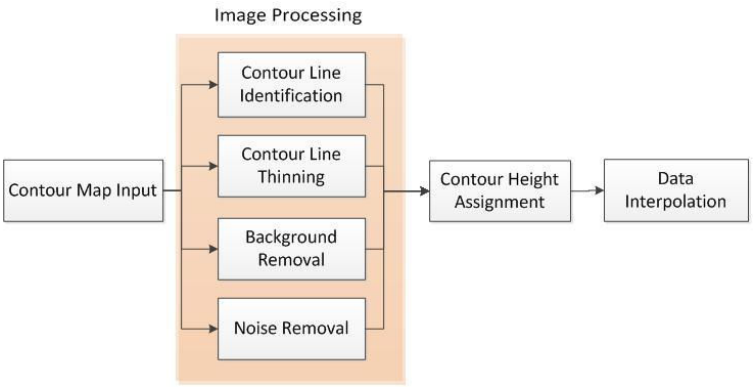


Figure 42: Digital elevation model (DEM) conversion process.

within the image are identified using the Canny edge detection algorithm. The Canny edge detection algorithm incorporates Gaussian noise reduction while also calculating diagonal gradient information to better handle noise sensitivity and erroneous pixel identification [64].

Although edge detection is a strong first step towards extracting contours, some contours within a typical 2D contour map are weighted (*i.e.*, thicker) than others. Typically, this is attributed to image quality or a deliberate attempt by the map originator to signify different elevation values. Regardless, following edge detection, a map with uniformly-weighted contours is required. As a result, we accomplish line thinning by repeatedly removing edges appearing at the output of the Canny algorithm and apply a threshold to assign edge-related pixels an intensity value of white (255) or black (0). Connectivity analysis and blob filtering help to remove any additional noise in the contour image upon completing the thinning process. The complete process is summarized in Algorithm 5.

5.1.3.2 From contour lines to elevation values

To automate the assignment of elevation values to identified contour lines, a three-phase approach is implemented. Following the line-thinning of the filtered image, 1's are assigned to pixels associated with each line contour and remaining pixels are assigned 0's. Next, our connected component labeling algorithm (line 13, Algorithm 5)

Algorithm 5 Contour identification and noise reduction

Require: input image X, number of iterations $N (N \geq 1)$, and threshold value $\alpha (0 < \alpha \geq 1)$

```
X ⇐ grayscale(X)
for i = 1 to N do
  C ⇐ Canny(X)
  X ⇐ X − C
  if X_{j,k} < Max(X) · α then
    X_{j,k} ⇐ 0
  else
    X_{j,k} ⇐ 255
  end if
  X ⇐ BlobRemove(X)
end for
X_{labeled} ⇐ Logical(X)
X_{labeled} ⇐ ConnCompLabelAlg(X)
return {X, X_{labeled}}
```

assigns a contour identification number to each uniquely grouped set of contour pixels. The identification step requires applying a pixel connectivity rule to differentiate between useful data pixels and noise from previous steps. Finally, the identification numbers are replaced with an analogous elevation height, organized prior to contour line assignment.

5.1.3.3 Data Interpolation

The remaining task in generating a three-dimensional terrain requires estimating the unassigned locations in the map space based on the assigned locations. For our interpolation scheme, we chose Delauny triangulation, since it does not require a model, responds well to discontinuities, and consistently generated the lowest RMS error values in preliminary testing [65]. When operating in continuous spaces with low standard deviation, the Delaunay triangulation interpolator provides a strong combination of computational efficiency and reduced RMS error in our experiments, both with spatial data acquired from simulation and hardware. To aid the linear Delaunay triangulation, we provide data at the boundary of the space for the purpose

of smoothing the generated terrain.

5.1.4 Validation of the DEM Maker

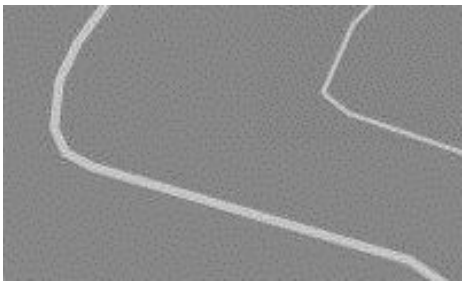
We, now, briefly summarize the details of our selected image processing algorithms used to convert a 2D contour map into a usable 3D DEM. We discuss how a small map segment is processed. Prior to comparison with any in situ data collected, we input the partial map segment into our system to confirm proper operation of the Canny edge detection algorithm. At the output of the Gaussian noise reduction, zero crossings are calculated (Equation 23).

$$\frac{\partial^2}{\partial n^2}\left(\mathbf{G} * \mathbf{I}\right) = \frac{\partial}{\partial n}\left(\frac{\partial \mathbf{G}}{\partial n} * \mathbf{I}\right) \tag{23}$$

$$\mathbf{G} = \sqrt{\mathbf{G}_x + \mathbf{G}_y}$$
$$\mathbf{\Theta} = \arctan\left(\frac{\mathbf{G}_y}{\mathbf{G}_x}\right)$$

Despite the specific features unique to our test segment, applying the Gaussian filter improves image quality significantly, removing pixelation and granularity, originally seen in Figure 43a. Following this, a single iteration of the contour thinning process, described earlier, at a specific threshold yields an improved contrast between contour and non-contour pixels (Figure 43b). After multiple passes of the thinning process, we obtain a more preferred image improvement (Figure 43c). We show a partial of the test segment after converting to a logical pattern of 1's and 0's (Figure 43d).

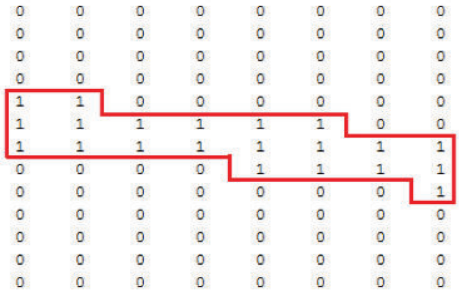
At the second stage of the map generation, the filtered image is converted to a series of 1's and 0's to mark contour pixels and background pixels, respectively. Our connected component labeling algorithm, then, properly labels each contour with its respective pre-assigned elevation value in preparation for interpolation. The third



(a) Smaller test contour map segment.

(b) Result of single pass thinning at 0.8 threshold.

(c) Result of 4-pass thinning at 0.8 threshold.



(d) Labeled contour (left) after use of connected component labeling algorithm.

stage (interpolation) was not necessary for such a small segment of data, and was applied to a more complete (*i.e.*, larger) map.

5.1.4.1 Benchmarking and Error Analysis

In confirming accuracy of the translation from 2D contour data to a 3D DEM, we compared our automatically-generated DEM with two 3D maps, each generated from the same 2D contour map (Figure 43) by a human manually extracting contours (considered an expert) and from data collected by a navigating robot. Acquiring data by these two alternative methods and producing 3D DEMs enabled multiple quantitative comparisons with our automated process. The surveyor robot executed a traditional parallel transect (or "lawnmower") pattern while collecting pitch and roll information. This data was then translated into a 3D map for this analysis using the same interpolation methods as were used at the output of the connected component labeling algorithm. A three-way comparison of data is presented at multiple lateral

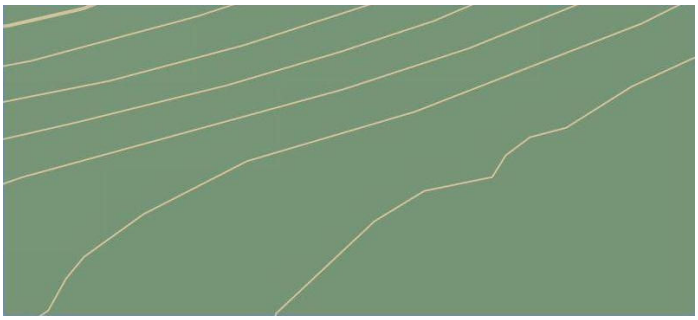


Figure 43: Complete contour map of test area.

Table 2: Error: Human expert-generated map vs. SECT-II map.

Resolution [m]	Relative (%)	Absolute [m]	RMSE [m]
0.2	20.175	0.184	0.234
1	23.983	0.210	0.280
2	33.237	0.311	0.390

spatial resolutions (Tables 2-4, Figure 44).

Part of our motivation for providing the tabulated error analysis is to emphasize the importance of obtaining high-resolution reference data points. Downscaling of our processed data causes a decrease in vertical accuracy and a less-reliable DEM. The more immediate purpose of our error data, however, is to highlight the benefit in automating the contour identification process. The consistent accuracy of the DEMs produced by our contour identification and thinning system should be noted in contrast to that of DEMs generated by the human expert. While map contours can be manually identified by hand, this is a time-consuming process. Additionally, there may exist subtle directional changes at the pixel level that can remain undetected

Table 3: Error: Algorithm-generated map vs. SECT-II map.

Resolution [m]	Relative (%)	Absolute [m]	RMSE [m]
0.2	20.256	0.185	0.234
1	24.012	0.210	0.279
2	33.234	0.310	0.388

Table 4: Error: Human expert-generated map vs. Algorithm-generated map.

Resolution [m]	Relative (%)	Absolute [m]	RMSE [m]
0.2	0.380	4.023e-3	6.066e-3
1	0.379	4.054e-3	5.761e-3
2	0.382	4.084e-3	5.974e-3

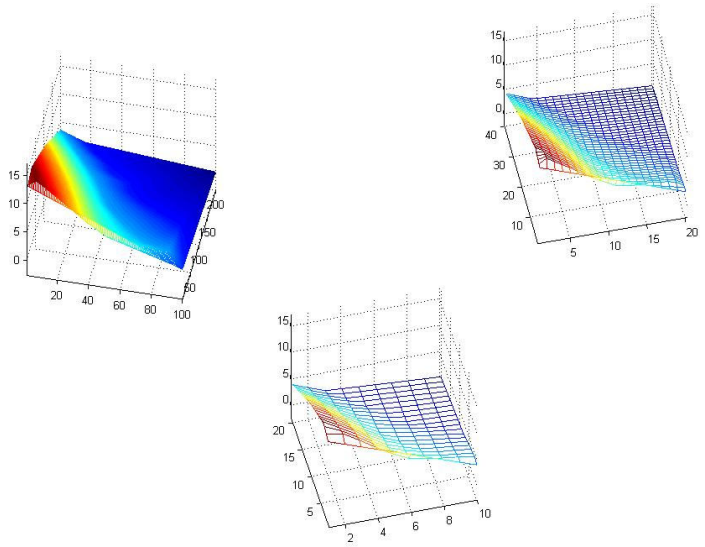


Figure 44: 3D-generated maps at various resolutions; 0.2 [m] (top-left), 1 [m] (top-right), 2 [m] (bottom-center)

Table 5: Average RMS vertical accuracy of 20 map samples at three horizontal resolutions of accuracy.

METHOD	1 [m] Res.	5 [m] Res.	10 [m] Res.
Cubic	1.234e-2	1.108e-2	9.789e-3
Linear	5.670e-3	5.370e-3	4.835e-3
Nearest	4.401e-2	3.828e-2	4.046e-2
v4	4.780e-3	4.922e-3	5.116e-3

by the human eye, yet, not to our system that considers a pixel-by-pixel analysis. Any differences in error between these methods (while on the order of 10e-3 [m]), are associated with border data assignment and the visual limitations of the human expert. An area of improvement for our system includes reducing residual noise from the application of multiple algorithms (*i.e.*, Canny edge and line thinning).

The DEM generated from data collected by our robotic agent serves as a secondary baseline against which we compare our automated DEM generation method. Again, our automated method shows strong agreement with the robot-based DEM, specifically, meeting ASPRS vertical accuracy standards (Class 1-3 map accuracy) for 0.2 [m] and 1 [m] horizontal resolutions [57].

To quantitatively understand the impact of different interpolation options in our DEM generation process, we collected and compared DEM data between the human expert and our algorithm for 20 different terrain examples accessed from the Atlanta Department of Watershed Management online map database (`http://gis.atlantaga.gov`). See Table 5. We notice a sensitivity to high-intensity clusters of pixels. This sensitivity remains despite the inclusion of our Gaussian noise reduction component. Regardless, the error between each method is extremely small, rendering the DEM generated by our automated method a strong approximation of the DEM from the human expert. The use of the Delaunay linear triangulation interpolation along with the v4 (or minimum curvature) method are more favorable

in DEM generation in contrast to cubic spline and nearest neighbor methods [66]. The terrains in our set of 20 are considered, by computer graphics literature, to be C_k continuous, where $k > 0$ [67]. This condition makes for a bounded terrain gradient, yielding less errors from spatial discontinuity, and therefore less error with the linear method. When k is increased, influencing the degree of spatial change, it is expected that Delaunay cubic spline interpolation would perform better than the other methods (*i.e.*, linear or v4).

When the impact of spatial resolution is considered, the Delaunay linear triangulation interpolator yielded an average decrease in error of 835 [mm] (row 2, Table 5). Unfortunately, this decrease comes as a result of lowering the spatial resolution (*i.e.*, increasing the distance between measurements by downscaling). In lieu of this observation, we suggest additional testing to identify any limits of spatial resolution at which two DEMs ought to be quantitatively be compared. Practically, contour maps are inherently inaccurate, especially for large spacings between contours (*i.e.*, spacings representing several feet in the vertical plane). The poor performance of the nearest neighbor interpolator is intuitively attributed to the absence of a sufficient number of data points near unidentified locations, leaving large gaps in the estimated spaces.

These simulated DEMs provide a ground truth to which we can apply our sampling methodologies (Section 4.1) and use as a reference for comparing with the estimated terrains resulting from those sampling strategies.

5.2 Validation of Contributions

When operating in large spaces, ideally, we would like to consider all possible combinations of sample configurations, as done in Section 3.2.1. Without sufficient computing resources, however, this exhaustive analysis becomes computationally intractable, especially for spaces where $MN > 16$ (Figure 12). Specifically, we consider a series

of spaces whose area, MN, is greater than 500 m^2. First, we evaluate the performance of the robotic navigation algorithms discussed in Chapter 4, contrasting the performance of our estimation methodology with single-trial Bernoulli estimation, evaluating our method's ability to estimate these larger spaces using each navigation option as a sampling mechanism (Contribution 1). Second, we will show how one of these navigation algorithms (lawnmower-informed) outperforms the others tested. (Contribution 2). Finally, we will evaluate the relative performance between each navigation option according to performance, *i.e.*, reconstruction error, and resource, *i.e.*, distance constraints (Contribution 3).

5.2.1 Informed Estimation Method: Contribution 1

We begin with the comparison between our informed estimation methodology and that of single-trial Bernoulli (hearafter referred to as "traditional") estimation. Each of the navigation algorithms discussed in Chapter 4 (lawnmower-traditional, lawnmower-random, and lawnmower-informed) are used as sampling schemes, providing unique sample configurations, which each estimation method uses to reconstruct the sampled space. In Chapter 3, we determined that our informed estimation method performed best on terrains spaces with low spatial frequencies ($\lambda_{Avg} < 25\%$). The theory behind our informed method also showed the resulting expected error defines a maximum limit on actual error generated for any space. With these observations in mind, we selected 100 unique terrain spaces, produced by our DEM maker software (Section 5.1.3), where the average spatial frequency is equal to 1.5 percent. Each DEM, originally defined in the continuous domain, is converted so that every location in the space is defined according to one of two classes from C (Section 3.1.1). Expected percent error is calculated according to Equation (24) while actual percent error is calculated using Equation 25, comparing the original DEMs, from which samples are collected, and the DEMs reconstructed from those samples.

$$\text{Percent Expected Error (\%)} = 100(\frac{E[error_{q_h}]}{MN}) \tag{24}$$

$$\text{Percent Actual Error (\%)} = 100(\frac{error_{q_h}^{Actual}}{MN}) \tag{25}$$

5.2.1.1 Applied Metrics to Simulated Terrain Models

We define our sample space, S, as a 115 [m] x 115 [m] area with 1 [m] spatial resolution, *i.e.*, $M = 115$ and $N = 115$. We randomized the locations of different undulating features (*e.g.*, hills and valleys) within each terrain based on the diamond-square (DS) algorithm used for generating fractal DEMs [60]. This is accomplished by changing the seed value of the random number generator used by the DS algorithm to designate the locations of local minimum and maximum values, and surrounding values between these extrema. Changing the seed value used in the DS algorithm ensures that no two terrains are identical. Once each terrain is generated, we quantize the value at each sample location to produce a series of quantized DEMs. Now, instead of collecting samples with information belonging to a continuous range of values, we will consider samples whose values are classified as belonging to one of the two classes in C (Section 3.1). Figure 45 shows one of the 100 simulated terrains generated and quantized for application of our sampling configurations.

We select a range of coverages over which we evaluate the performance of each navigation algorithm. These coverages are defined as a function of the ratio between the number of samples collected, B, and the number of total possible samples in S, MN=(115x115). For a specific measure of percent coverage, each algorithm collects B samples.

$$B = \begin{cases} (M)(\text{Number of swaths}) & \text{if navigating west to east/east to west} \\ (N)(\text{Number of swaths}) & \text{if navigating south to north/north to south} \end{cases} \tag{26}$$

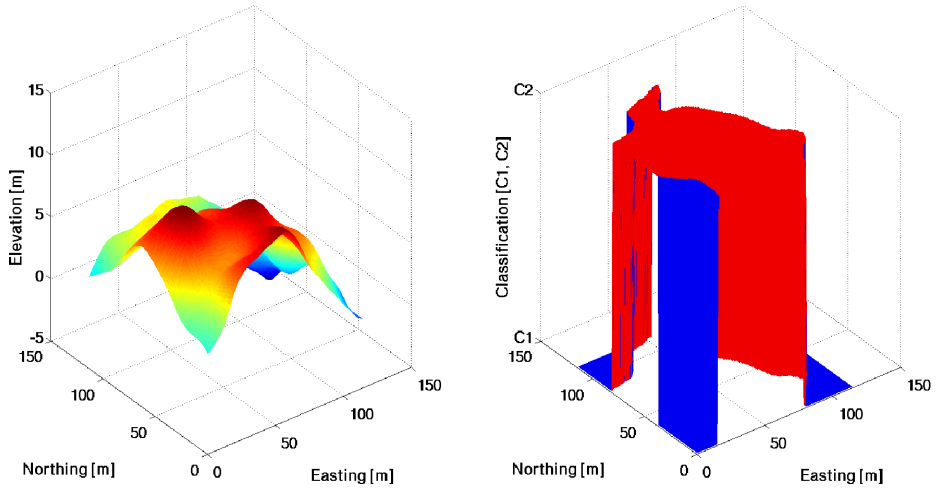


Figure 45: Example of a randomly generated DEM in continuous space (left) and corresponding DEM in quantized space (right) used for validation of navigation algorithms.

As a quantity, B is a constant multiple of a specific number of swaths and each swath contains M or N samples, depending on the direction of navigation, *i.e.*, west to east or south to north (Table 6).

Table 6: Coverage ranges for 115 [m] x 115 [m] test area at meter resolution.

Number of Swaths	Sample Quantity, B	Coverage (%)
2	230	1.739
3	345	2.609
4	460	3.478
5	575	4.348
6	690	5.217
7	805	6.087
8	920	6.956
9	1035	7.826
10	1150	8.696
11	1265	9.565
12	1380	10.435
14	1610	12.174
16	1840	13.913
18	2070	15.652
22	2530	19.130
28	3220	24.348
37	4255	32.174

Similar to the results from Chapter 3, we validate the performance of our informed estimation methodology against the traditional estimation method. Instead of averaging the performance of all Q_B sampling configurations for the test space, S, instead, we apply the configurations generated by the navigation algorithms discussed in the previous chapter. For a designated number of reference swaths (or coverage given B samples, per Table 6), lawnmower-traditional, lawnmower-random, and lawnmower-informed algorithms each generate a unique set of sampling configurations. We use the metrics of average expected error and average actual error, first introduced in Section 3.2, to evaluate the performance trends that exist over a specific range of coverage. First, we consider the performance of our informed estimation method using samples collected according to the lawnmower-traditional navigation strategy (Figure 46).

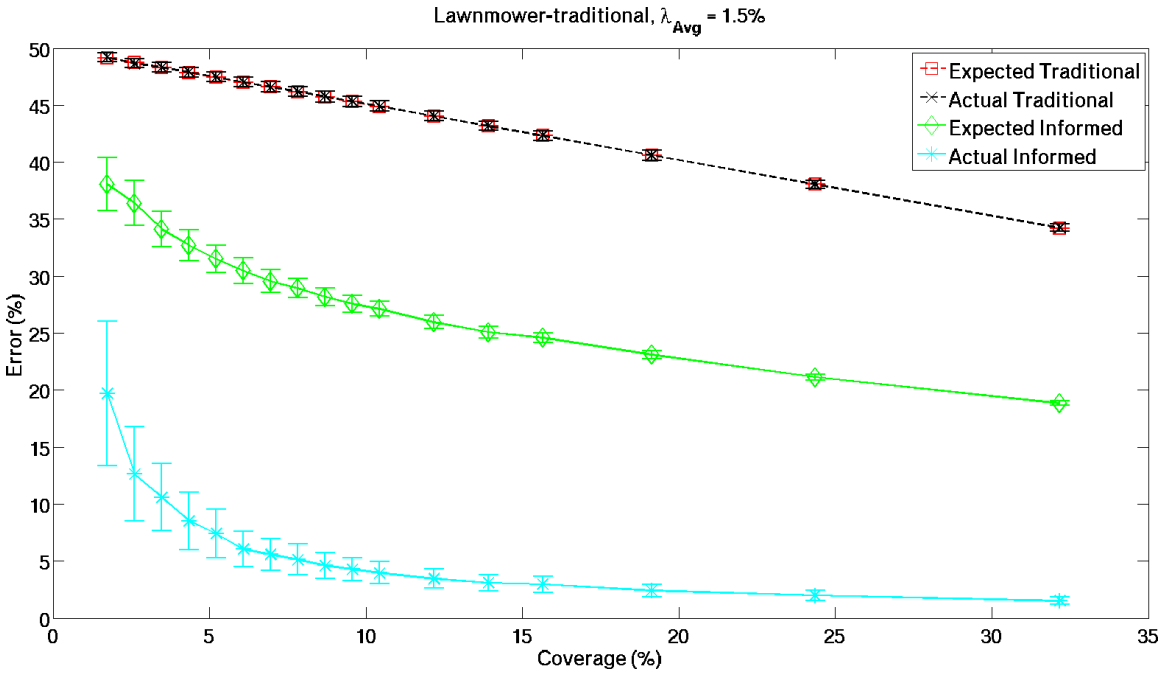


Figure 46: Performance of informed estimation using lawnmower-traditional navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=1.5%.

The error trends (expected and actual) in Figure 46 confirm the theoretical performance of our estimation method, demonstrated in Chapter 3 (Section 3.2.2). Specifically, the expected and actual error generated by our informed estimation outperforms that of the traditional estimation method. Furthermore, the expected error calculated by our method produces an upper bound on the actual error generated. By using this expected error as an upper bound, a scientist is guaranteed a specific cap on the actual error produced when using our informed method to reconstruct information in the space. Using our informed method to approximate an upper bound of actual error is significant particularly when professional standards relating to reconstructed map accuracy must be used. Knowing an upper error bound is analogous to the application of ASPRS standards for contour map accuracy.

Additionally, there exists a nonlinear trend in both the expected and actual error of the informed method. The "knee" in this data allows us to identify the amount of coverage at which the rate of improvement over traditional estimation is greatest. This is important when considering what coverage amount, when combined with our estimation method, will generate the greatest performance, *i.e.*, lowest expected error, relative to the traditional estimation method. In Chapter 3, we approximated this coverage value as 30-40 percent of the number of samples in S (Section 3.2.2). Here, based on a much larger space, we identify a lower coverage range of 5-10 percent.

We repeat analysis for the expected and actual error generated from our informed estimation based on samples collected according to lawnmower-random navigation (Figure 47)

Again, as with the trends for lawnmower-traditional, we observe in Figure 47 the capability of our informed estimation method to accomplish two tasks. The first task is successfully generating actual error that falls below the actual error produced by traditional estimation. The second task is producing an expectation of error that approximates an upper bound for actual error when the spatial frequency of the test

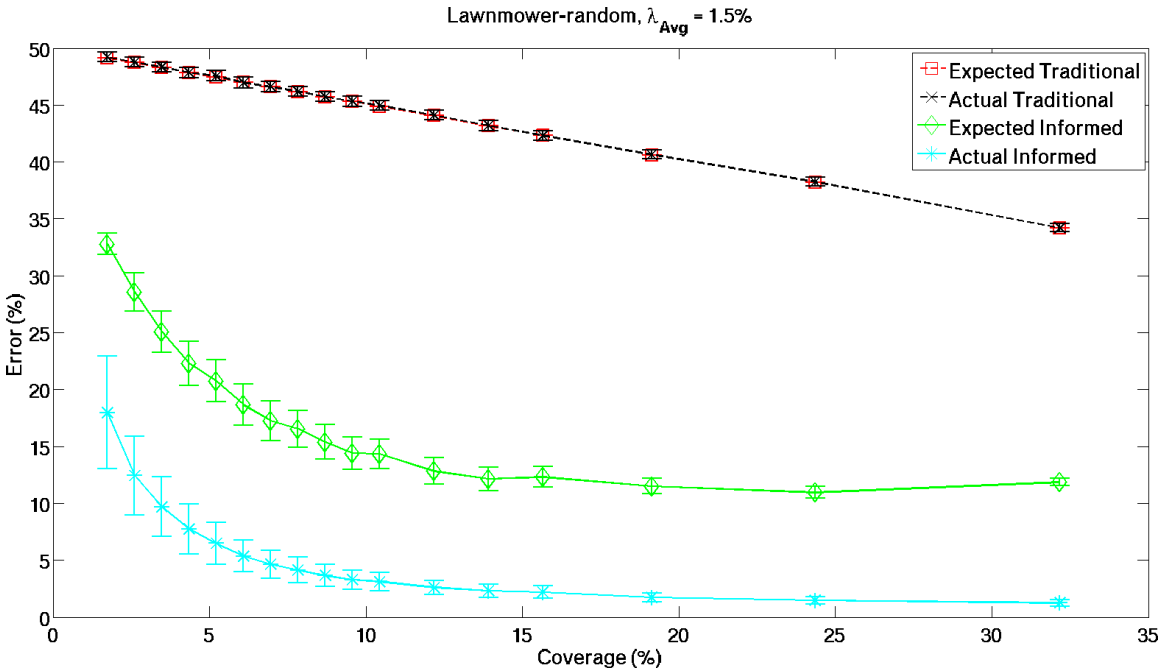


Figure 47: Performance of informed estimation using lawnmower-random navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=1.5%.

area, λ_{Avg}, is equal to 1.5 percent. We also notice the range of coverage, where the rate of improved error for the informed method over traditional estimation is greatest, as 5-10 percent. We consider one more series of performance trends based on the samples collected according to lawnmower-informed navigation (Figure 48).

The observations made for the trends in Figure 48 mirror those mentioned already for Figures 46 and 47.

Figures 46-48 illustrate the ability of our informed estimation method to approximate a set of dual-class DEMs with actual error lower than the expected value. This finding is specifically accurate for terrain examples whose spatial frequency, λ_{Avg}, is less than 25 percent. In the Appendix, we discuss the impact of spatial frequency identified in our validation, confirming original theoretical observations made in Chapter 3 (Appendix B).

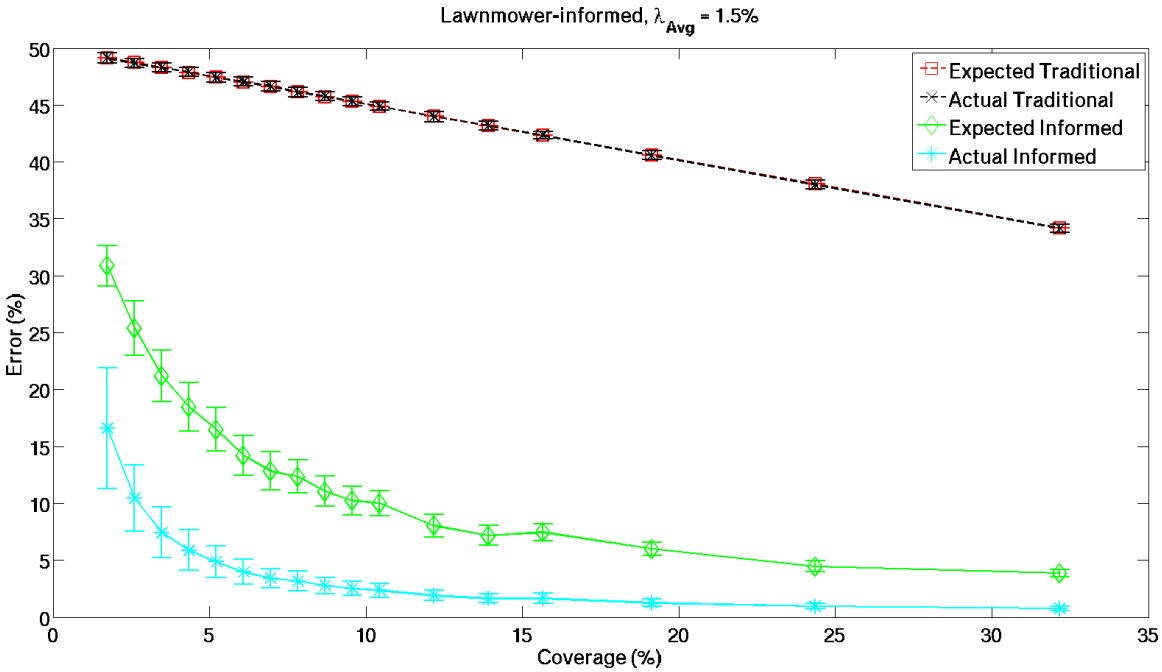


Figure 48: Performance of informed estimation using lawnmower-informed navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=1.5%.

5.2.2 Constraint Comparison of Informed Navigation Algorithm: Contributions 2 and 3

We discuss the relative performance between navigation algorithms. Given that 100 DEMs were included for each spatial frequency, this condition produces 100 unique configurations, each with a measure of performance and required resources, *i.e.*, actual error and distance (Section 3.4.1), respectively. Now, with our restriction on spatial frequency, we contrast the unique sample configurations generated by each of the navigation algorithms tested. We selected lawnmower-traditional (*i.e.*, raster scanning) as a baseline for its "uninformed" trait as a navigation strategy, since it does not require any input beyond area boundary information for it to be applied to a search space. The other similar navigation pattern selected as a baseline for comparison is termed lawnmower-random. Lawnmower-random provides a similar structure as lawnmower-traditional, but designates sampling locations as waypoints that are randomly distributed but centered around each reference swath applied. This is most

closely related to the simple random sampling method (Sections 2.1 and 2.4).

We presented our measure of a sampling configuration's required resources in terms of a raw value, D_{Avg}^{NN} (Chapter 3). Here, we compare between different sample configurations' resources as a percentage. This percentage is specifically calculated as the ratio of D_{Avg}^{NN} to a nearest neighbor upper limit distance, D_{BW}^{NN}. The value, D_{BW}^{NN}, is a function of the swath width between reference swaths used during navigation (Equation 27).

$$D_{BW}^{NN} = \sqrt{\delta^2 + (\text{Swath Bandwidth})^2} \tag{27}$$

$$\text{Resources } (\%) = 100(\frac{D_{Avg}^{NN}}{D_{BW}^{NN}}) \tag{28}$$

We normalize D_{Avg}^{NN} using this swath width since D_{BW}^{NN} represents the farthest nearest-neighbor distance possible between two individual samples when B total samples are collected (Figure 49). The reader may also refer to Figures 29, 30, and 39 for a visual depiction of how these swath widths change with increasing coverage.

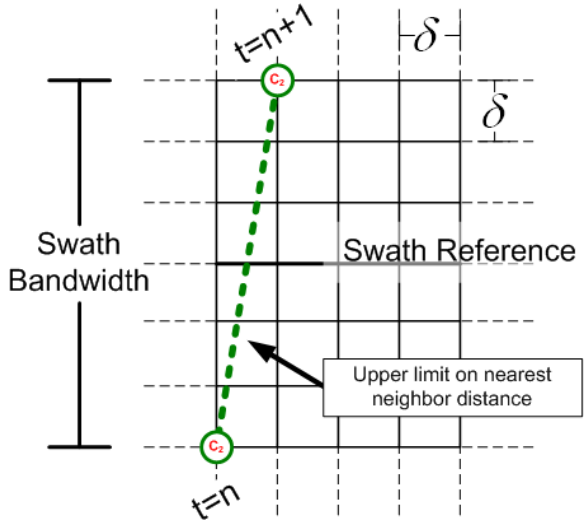


Figure 49: Changing maximum nearest neighbor distance, D_{Max}^{NN}, used to calculate the average resources incurred by a specific navigation trajectory.

We highlight the performance of our estimation-based lawnmower-informed navigation scheme to demonstrate its ability to achieve desirable performance (low reconstruction error) while conservatively using resources (minimal distance traversed) to accomplish the sampling task. Moreover, this navigation scheme accomplishes these goals by using our informed estimation methodology as its input for sample selection decisions. For these tests contrasting performance and resources, we hold the spatial frequency of the terrains tested constant at 1.5 percent, considering multiple coverages. Each data point within each cluster of Figure 50 represents a unique sampling configuration obtained by a specific navigation algorithm. We make the comparison between each navigation algorithm relative to our two metrics of actual performance and resources on the y and x axes, respectively (Figures 51).

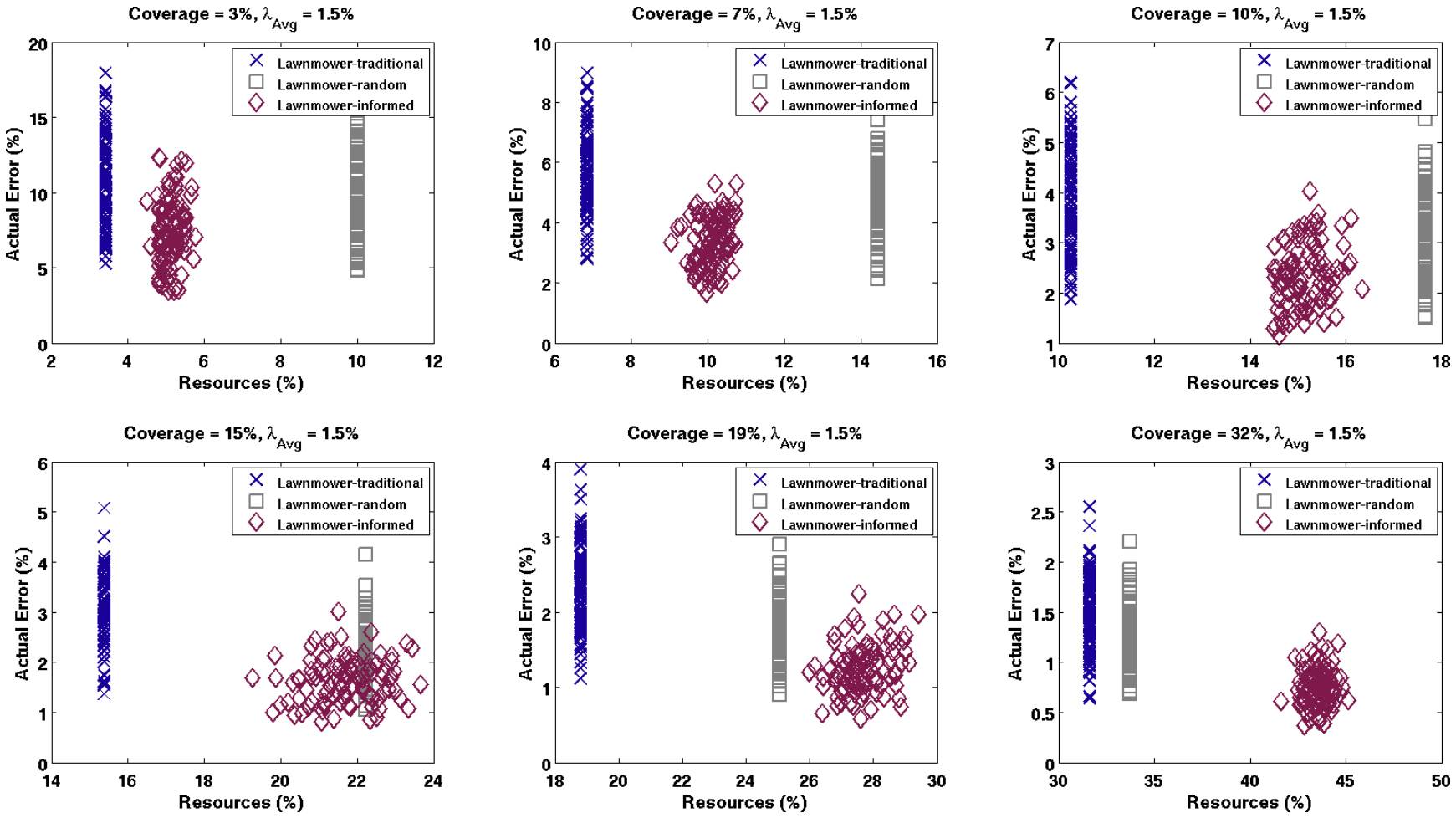


Figure 50: Comparison between navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed, as a function of performance (actual error) and resource (distance) metrics for multiple percent coverages and a specific spatial frequency, λ_{Avg}=1.5%.

For each measure of coverage, the distribution of actual errors for the lawnmower-informed navigation is positioned lower along the y-axis than those for lawnmower-traditional and lawnmower-random, showing an improvement in actual error across the 100 terrains tested. Additionally, for coverages less than 15 percent, our proposed lawnmower-informed navigation method outperforms lawnmower-random in terms of average nearest neighbor distance required to travel between samples, indicating an improvement in resource usage. A more quantitative comparison can be made using the data presented in Table 9.

Table 7: Relevant performance and resource data for navigation strategies lawnmower-traditional, lawnmower-random, and lawnmower-informed navigation for specific coverages presented in Figure 50.

	Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)	Avg. Resources (%)
Lawnmower-traditional		12.622	4.178	5.331	2.702	22.957	2.702	2.702
Lawnmower-random	2.563	12.425	3.476	5.837	9.068	19.977	9.068	9.068
Lawnmower-informed		10.438	2.917	4.423	3.821	18.442	4.413	4.085
Lawnmower-traditional		5.585	1.392	2.790	7.125	8.975	7.125	7.125
Lawnmower-random	6.836	4.642	1.243	2.132	15.082	8.015	15.082	15.082
Lawnmower-informed		3.401	0.830	1.671	10.405	5.293	10.922	10.556
Lawnmower-traditional		3.960	0.962	1.868	11.043	6.193	11.043	11.043
Lawnmower-random	10.253	3.111	0.812	1.490	19.028	5.459	19.028	19.028
Lawnmower-informed		2.354	0.621	1.142	15.748	4.023	16.442	16.323
Lawnmower-traditional		2.956	0.702	1.361	16.440	5.074	16.440	16.440
Lawnmower-random	15.380	2.197	0.554	1.051	23.944	4.144	23.944	23.944
Lawnmower-informed		1.641	0.441	0.809	22.533	3.009	22.999	23.174
Lawnmower-traditional		2.412	0.554	1.119	19.612	3.894	19.612	19.612
Lawnmower-random	18.798	1.734	0.393	0.900	26.144	2.896	26.144	26.144
Lawnmower-informed		1.257	0.319	0.575	28.803	2.238	28.753	28.881
Lawnmower-traditional		1.519	0.347	0.643	31.623	2.548	31.623	31.623
Lawnmower-random	31.614	1.253	0.285	0.681	33.765	2.200	33.765	33.765
Lawnmower-informed		0.756	0.181	0.371	42.888	1.301	43.632	43.600

We also present similar plots of performance (expected error) and resources along the y and x axes, respectively (Figure 50).

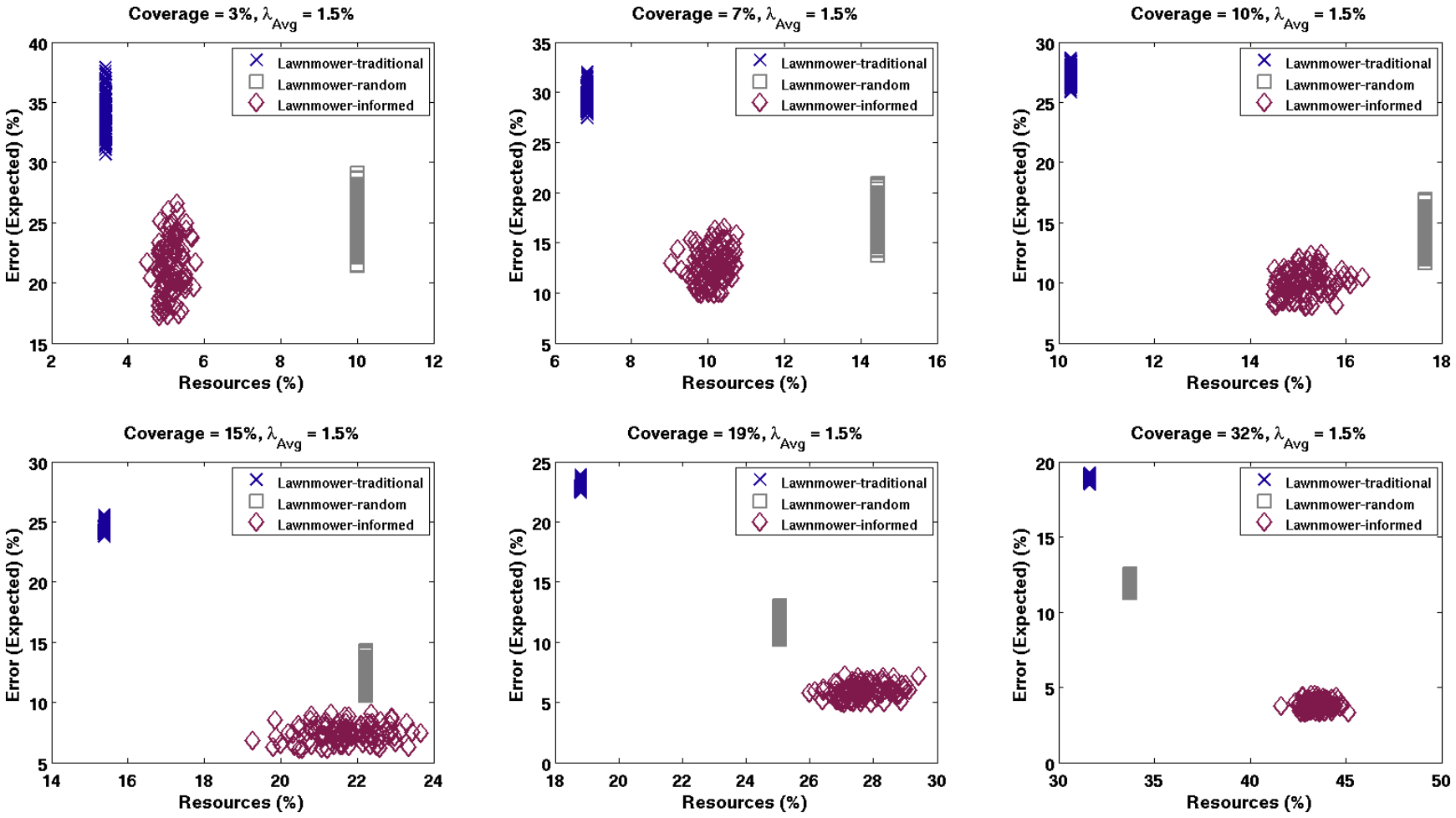


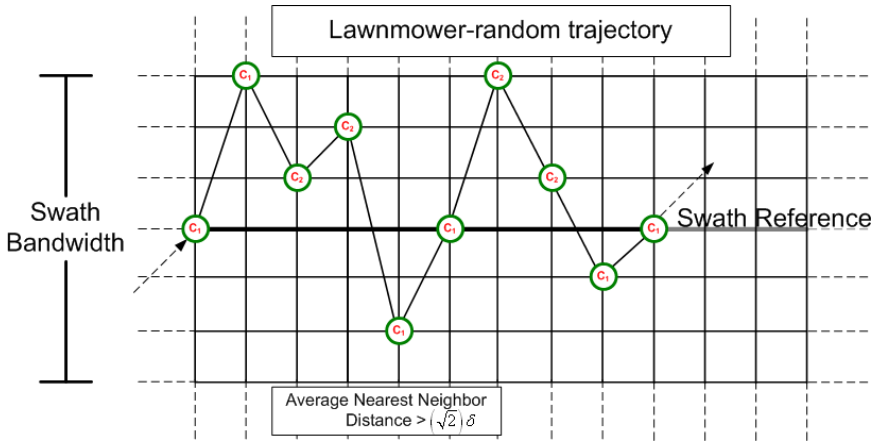
Figure 51: Comparison between navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed, as a function of performance (expected error) and resource (distance) metrics for multiple percent coverages and a specific spatial frequency, λ_{Avg}=1.5%.

We present the results of expected performance versus resources to confirm the usefulness of our informed estimation method and its ability to generate expected error that acts as an upper limit to the actual error generated. From the plot in Figure 51, the scale along the y axis includes values much greater than that found in Figure 50, consistent with our earlier observations, as originally discussed in Section 5.2.

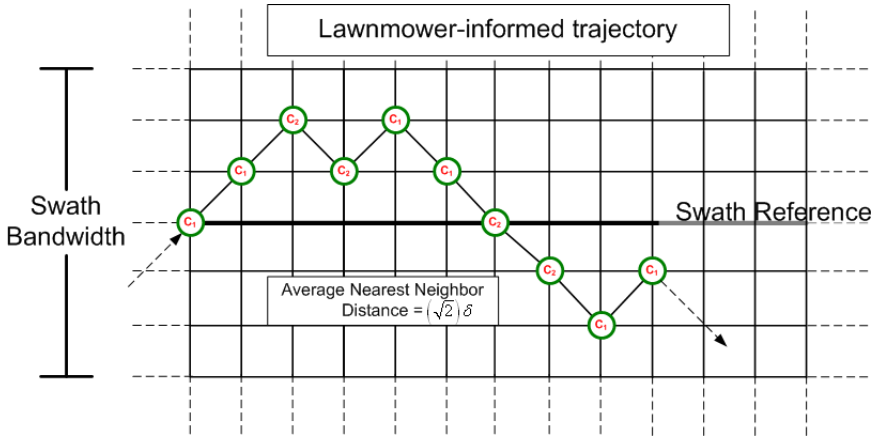
For these comparisons between navigation schemes, as coverage increases, we observe a decrease in the resources required by lawnmower-random versus that required by lawnmower-informed. This decrease in average nearest neighbor distance is attributed to both the decreasing bandwidth between reference swaths and the variety of sample selection that occurs as coverage, and subsequently swaths, increases (Figures 52-53).

As the swath bandwidth is decreased, due to an increase in sampling/coverage, the navigating agent experiences a spatial restriction in the number of candidate samples that may be collected along each swath. As a result, lawnmower-random navigation will select more samples of a closer distance, *i.e.*, δ, more frequently than will lawnmower-informed. This selection of closer samples lowers the overall amount of resources incurred by lawnmower-random for high coverages. Lawnmower-informed, however, is designed to select subsequent samples that provide the best combination of reduced estimation confidence (η_{Avg}) and reduced proximity to the agent's current location. While proximity is an important criteria in the sample selection, it is secondary to the criteria of reduced estimation confidence, therefore, its trajectory will seldom select samples immediately adjacent, *i.e.*, positions located less than $(\sqrt{2})\delta$ away from the agent's current position. The behavior of these navigation algorithms is described in greater detail in Chapter 4.

Under the conditions tested in Figure 50, it is observed that lawnmower-random,



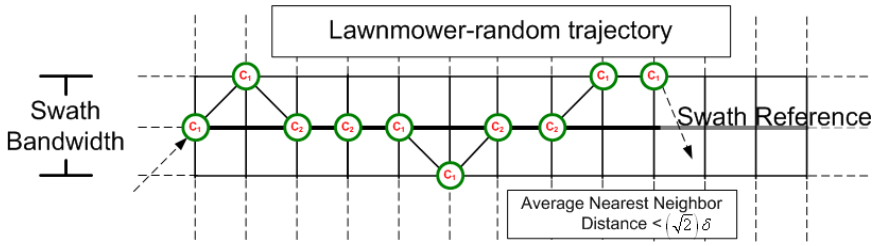
(a) Lawnmower-random trajectory.



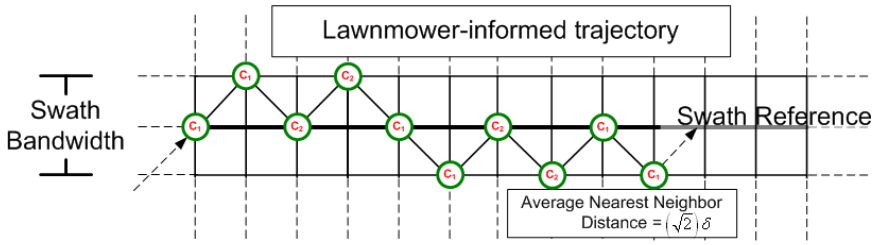
(b) Lawnmower-informed trajectory.

Figure 52: Comparison of average nearest neighbor distances incurred by lawnmower-random and lawnmower-informed navigation algorithms for low coverage, *i.e.*, large swath bandwidth.

on average, generates actual and expected error lower than that of the other two navigation options. Meanwhile, lawnmower-random also requires the greatest amount of resources for coverages less than 15 percent. In contrast, the lawnmower-informed navigation produces configurations that provide the best combination of lowest error (expected or actual) and minimal resource usage. The distribution of error associated with lawnmower-traditional navigation is greater than that of lawnmower-informed but with significantly lower resources. It is for this reason that we recommend application of Equation (21). A scientist can define their own design-specific constraints



(a) Lawnmower-random trajectory.



(b) Lawnmower-informed trajectory.

Figure 53: Comparison of average nearest neighbor distances incurred by lawnmower-random and lawnmower-informed navigation algorithms for high coverage, *i.e.*, small swath bandwidth.

(α_P, α_R, $error_{Max}$ and D^{NN}_{Max}) to determine, for example, whether the cost of applying lawnmower-traditional navigation with their robotic survey system outweighs the benefit of improving performance by applying lawnmower-informed navigation. The data presented in Figures 50 and 51 embodies this added value of analysis. We make note that, for all three sets of configurations, when increasing coverage, there exists a corresponding decreasing in error and increase in required resource percentage. This is seen in both Figures 50 and 51.

5.2.2.1 Performance of Informed Navigation on Simulated DEMs for Continuous Spaces

Our last validation includes applying the trajectories of each navigation method, as acquired from navigation within a quantized space, to a set of continuously-valued DEMs, as produced by our DEM maker (Figures 29, 30, and 39). When we apply the sampling locations for these navigation schemes to a continuous space, then estimate the continuous values at these locations using a popular 3D interpolator (*i.e.*, quadratic interpolation) to reconstruct a 3D map, we are able to generate a measure

of RMS error for each sampling configuration. Using the ASPRS standards outlined (Section 5.1.2), we gain valuable insight into how the sampling configurations generated by our lawnmower-informed navigation, although a function of quantized data, perform in a continuous space (Figure 54).

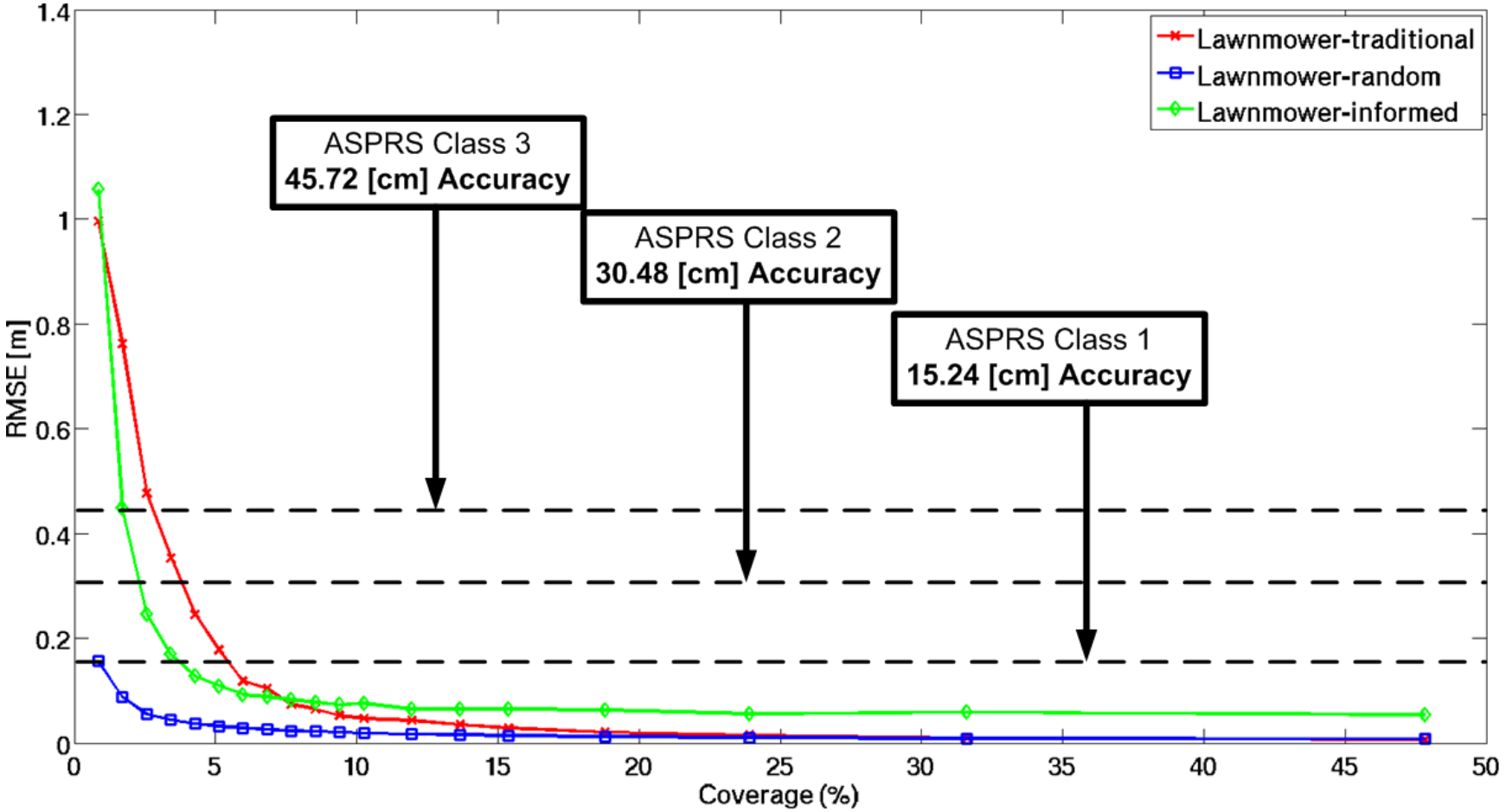


Figure 54: Comparison of RMS errors, produced by navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed to ASPRS Class 1, 2, and 3 mapping standards, as a function of coverage in a continuous domain (simulated DEMs).

We notice from the RMS error produced that for coverages between 2 and 7 percent, the lawnmower-informed navigation generates reconstruction error that is improved over that of lawnmower-traditional. Additionally, using lawnmower-informed as a sampling strategy will produce DEMs that meet ASPRS Class 1, 2 and 3 map accuracy standards for coverages greater than 4, 3, and 2 percent coverage, respectively. Lawnmower-traditional and lawnmower-random also generate reconstruction errors that meet these ASPRS mapping standards, but we see, from Figure 54, the potential of using the information sampled in situ to drive the navigation decisions. We discuss the observation more as a part of our recommended future work (Section 6.2).

5.3 Supplemental Validation and Observations

5.3.1 Applied Metrics to Realistic Terrain Models

Simulated terrains are helpful proof-of-concept tools for testing and evaluation, yet more realistic terrain models are necessary to ensure the consistent performance of our informed estimation methodology and associated sampling techniques. A set of 20 DEMs were extracted from publically available contour map data to continue testing the effectiveness of these algorithms [68]. An example DEM and its quantized counterpart DEM, shown in Figure 55, reflect characterisitics found in most of the DEMs in this test set.

Unlike the simulated terrain DEMs from Section 5.2.1.1, the size of the first set of terrains identified for extraction was, on average, representative of a 20 [m] x 10 [m] area with centimeter resolution (*i.e.*, $\delta < 0.1$ [m]). As a consequence of this increase in resolution, there exists significantly more sample locations within the search space. The range of coverages is displayed in Table 8.

An additional set of 10 terrain maps, representing soil moisture data as collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument

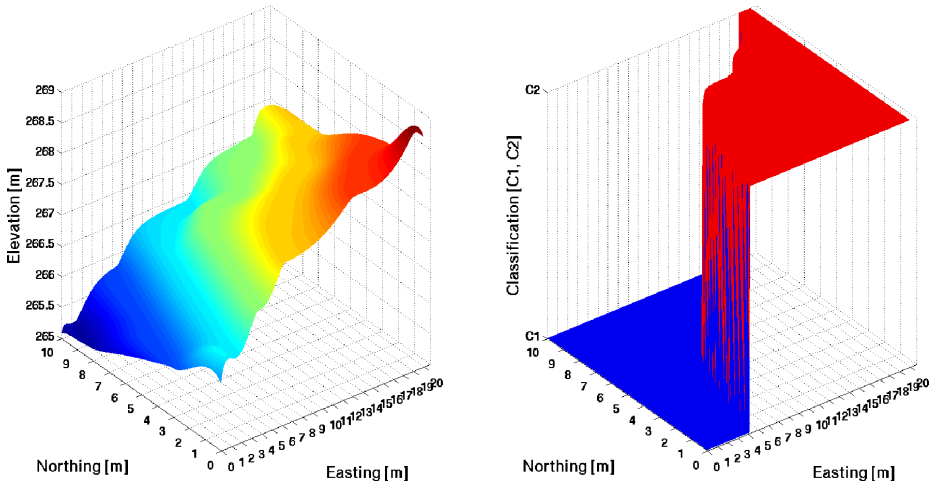


Figure 55: Example of a DEM extracted from 2D contour image data of a real terrain sample (left) and its quantized version (right).

Table 8: Coverage ranges for 20 [m] x 10 [m] test area at centimeter resolution.

Number of Swaths	Sample Quantity, B	Coverage (%)
1	485	0.658
2	970	1.316
3	1455	1.974
4	1940	2.632
5	2425	3.289
6	2910	3.947
7	3395	4.605
8	3880	5.263
9	4365	5.921
10	4850	6.579
11	5335	7.237
12	5820	7.895
13	6305	8.553
14	6790	9.211
16	7760	10.526
18	8730	11.842
21	10185	13.816
24	11640	15.789
29	14065	19.079
37	17945	24.342

(aboard Terra and Aqua satellites), was retrieved from an online database (`http:`

//modis.gsfc.nasa.gov/data/dataprod/). This data was obtained to conduct further testing of our estimation method and the relative performance of our navigation strategies, but for quantized data with a greater average spatial frequency (approximately 5 percent), since that of the translated contour map data is very small (λ_{Avg}=0.25%). We also include this MODIS data to demonstrate that our work can be applied regardless of the nature of the environmental phenomena sampled and estimated across an area, *e.g.*, elevation, soil moisture, or even radiation. This second set of data is more analogous, in size and resolution, to the simulated DEMs from Section 5.2.1.1, each representing a 115 [m] x 115 [m] area with meter resolution. As a result, we refer the reader to Table 6 for a range of reference swaths and their associated coverages.

Just as shown in Section 5.2.1.1, we applied sampling configurations generated by the different navigation algorithms from Chapter 4 to this new set of data maps. We evaluated the performance of our estimation method in contrast to traditional estimation according to established metrics, *i.e.*, average percent error (expected and actual). First, we tested the impact of sampling configurations, generated from the lawnmower-traditional navigation scheme, on these two sets of data, each set exhibiting an average spatial frequency equal to 0.25 percent and 5 percent (Figures 56-57).

We can confirm, based on Figures 56 and 57, the similarities in performance trends of expected and actual error tested on realistic data to those seen for simulated terrains with low spatial frequency, *i.e.*, 1.5 percent. First, our informed estimation method calculates expected error that exceeds its actual error and serves as an upper boundary for all coverages tested. Second, the actual error generated by the informed estimation method outperforms the actual error resulting from reconstructing a space with the traditional estimation. For data sets with both a spatial frequency of 0.25

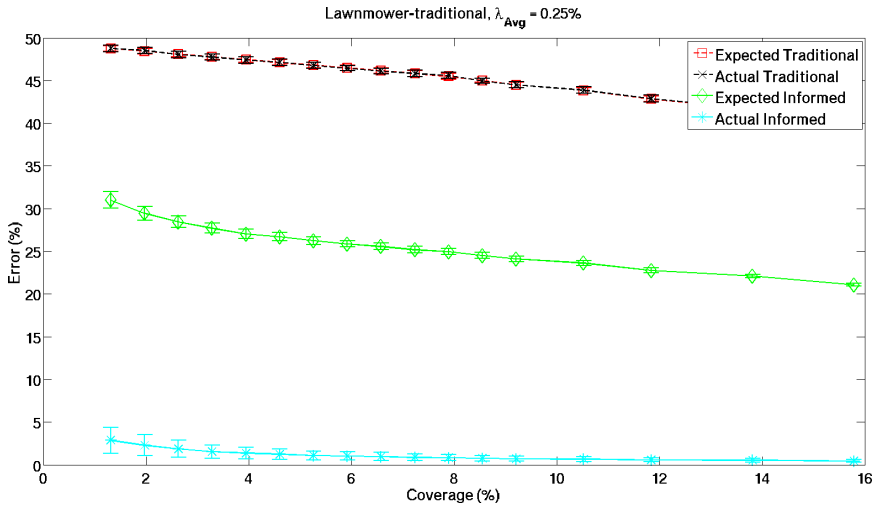


Figure 56: Performance of lawnmower-traditional navigation as a sampling strategy applied to DEMs extracted from contour map data at λ_{Avg}=0.25%.

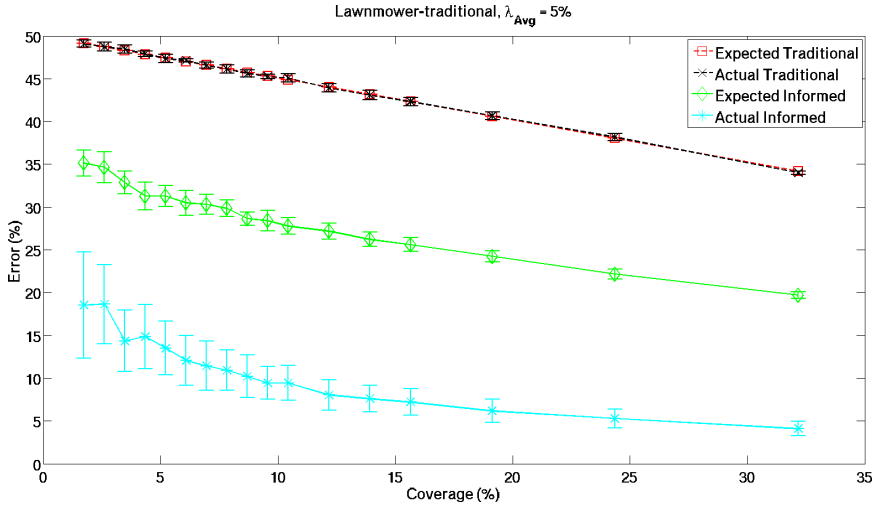


Figure 57: Performance of lawnmower-traditional navigation as a sampling strategy applied to DEMs extracted from MODIS map data at λ_{Avg}=5%.

and 5 percent, the actual error of the informed method outperforms traditional estimation by more than 30 percent. We also make mention of the coverage amount at which the maximum improvement of the informed method over traditional estimation takes place. Specifically, for data sets with $\lambda_{Avg} = 0.25\%$, because of the centimeter resolution of the space, there exist nearly five times as many samples in the space, S, than in our simulated set of DEMs. Therefore, we identify a slight reduction in

the coverage range at which the informed estimation method maximizes its improvement over traditional estimation as approximately 3-5% (Figure 56). This range is in contrast to that observed in Section 5.2.1.1 (approximately 5-10%). Since, for our data set extracted from MODIS data, the size and spatial resolution of all 10 terrain examples is the same as that of our simulated DEMs, the coverage range associated with maximum improvement of informed estimation over traditional is also approximately 5-10% (Figure 57). Both of these coverage ranges are observed by locating the "knee" of the exponential trend in expected and actual error for the informed method.

These results are also generated for sampling configurations produced by lawnmower-random navigation (Figures 58-59).

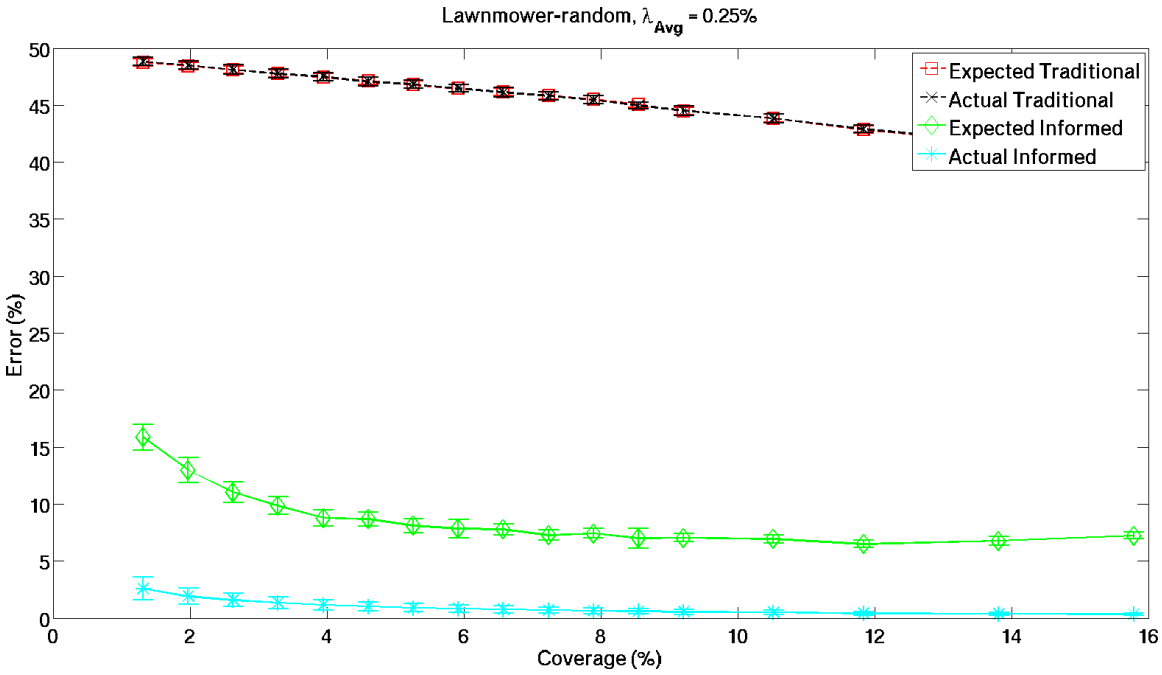


Figure 58: Performance of lawnmower-random navigation as a sampling strategy applied to DEMs extracted from contour map data at λ_{Avg}=0.25%.

We notice, for data maps of both spatial frequencies (0.25 and 5 percent), a closer approximation of actual error by the expected error trend in Figures 58 and 59. This

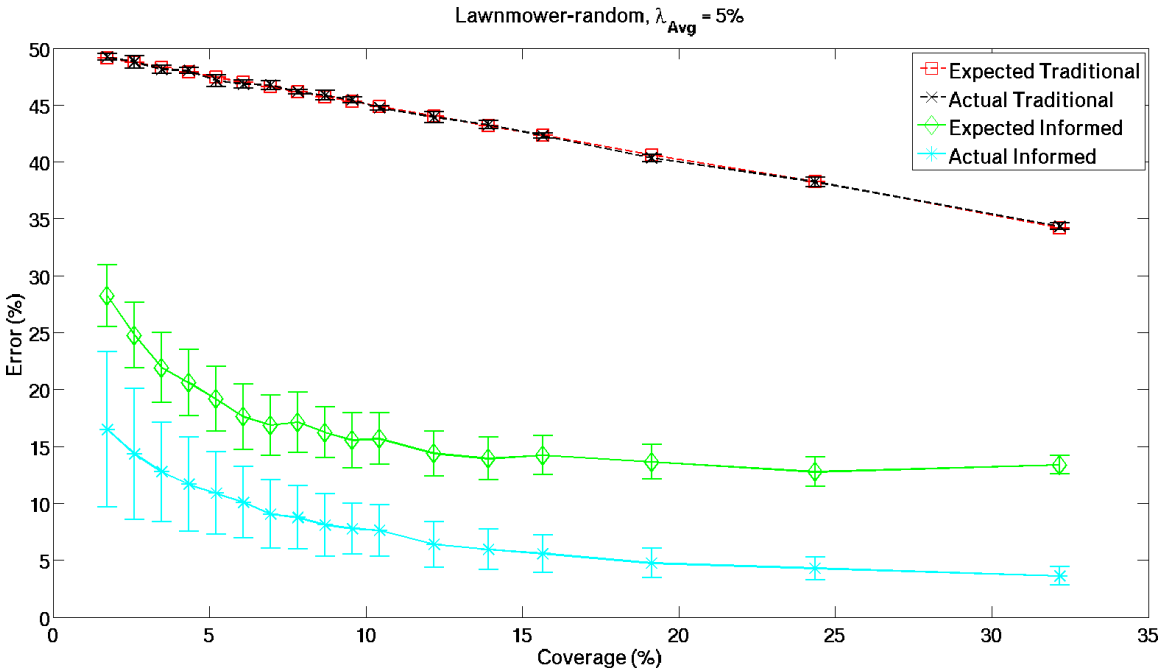


Figure 59: Performance of lawnmower-random navigation as a sampling strategy applied to DEMs extracted from MODIS map data at λ_{Avg}=5%.

observation is attributed to the sampling configurations generated by lawnmower-random and confirms that both the estimation method and the spatial relevance of selected samples contribute to the overall performance metrics. This observation is in addition to the improvement of our informed estimation method over traditional estimation and the application of the expected error trend as an upper bound guarentee on actual error produced by our method.

We conclude discussion on this set of data by including the performance of sampling configurations inspired by lawnmower-informed navigation (Figures 60-61).

The performance trends in Figures 60 and 61 more closely resemble those of lawnmower-random sampling. The similarity is with respect to the smaller gap between expected and actual error trends generated by our informed estimation method. This similarity is to be expected when comparing the improved spatial distribution of these two navigation schemes' trajectories (Figures 30 and 39), *i.e.*, there is strong

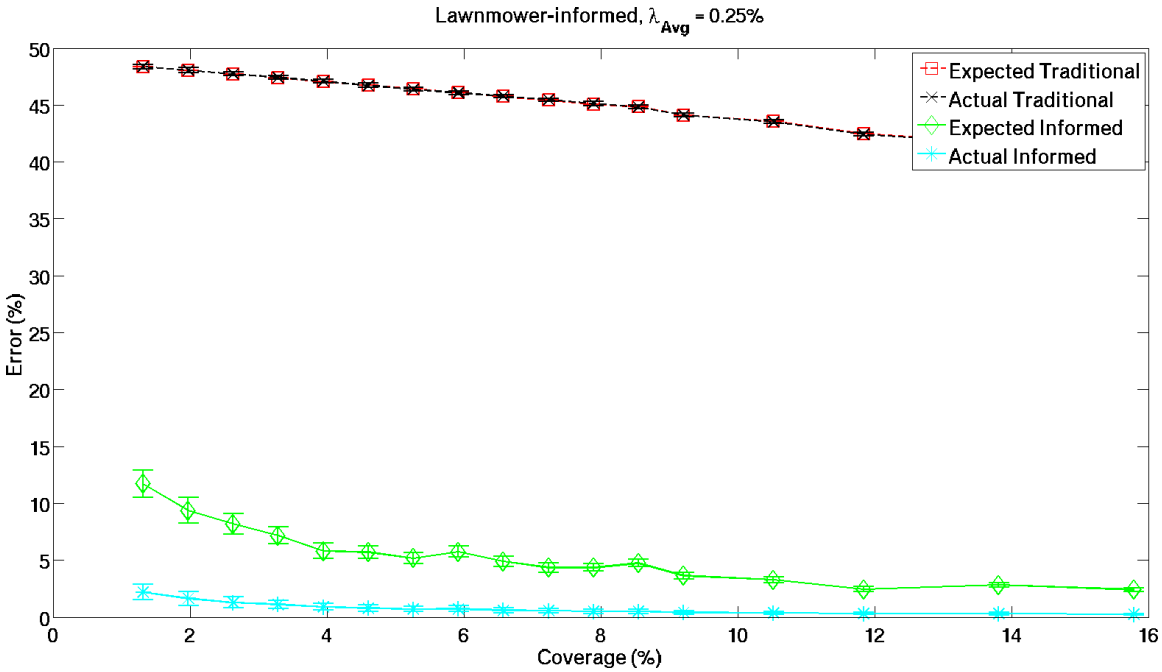


Figure 60: Performance of lawnmower-informed navigation as a sampling strategy applied to DEMs extracted from contour map data at λ_{Avg}=0.25%.

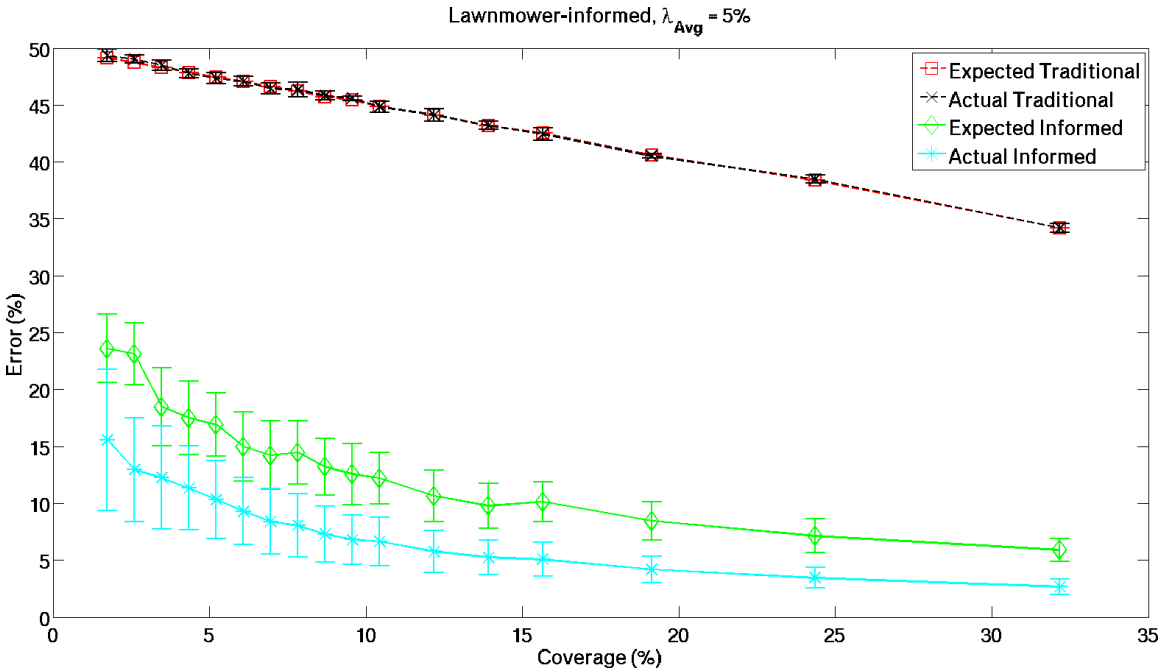


Figure 61: Performance of lawnmower-informed navigation as a sampling strategy applied to DEMs extracted from MODIS map data at λ_{Avg}=5%.

agreement in the spatial diversity of samples collected according to both lawnmower-random and lawnmower-informed navigation. Our informed method continues to

generate an expected error trend usable as an upper bound on actual error, but with less flexibility for actual error to vary, *i.e.*, a smaller gap between the two trends. For all three sets of navigation-inspired sampling configurations, trends of actual error fall below 5 percent using our informed method on DEMs with a spatial frequency of 0.25 percent at coverages of less than 2 percent. This is in contrast to the required minimal coverage of 6 percent or more to generate actual error trends less than 5 percent for simulated DEMs at a spatial frequency of 1.5 percent (Figures 46-48).

5.3.1.1 Comparison between navigation algorithms: Constraint analysis

Next, we present a visualization of how configurations associated with each of our navigation algorithms are distributed according to the two metrics of performance and resource usage. Although this relative comparison of which navigation option is the "best" is a function of user-defined constraints (Section 3.4), we make observations that facilitate a qualitative comparison. For a range of coverages, we present the average performance of each of our navigation algorithms, used as sampling schemes, as a function of their required resources for our set of 20 DEMs with an average spatial frequency of 0.25 percent (Figure 62).

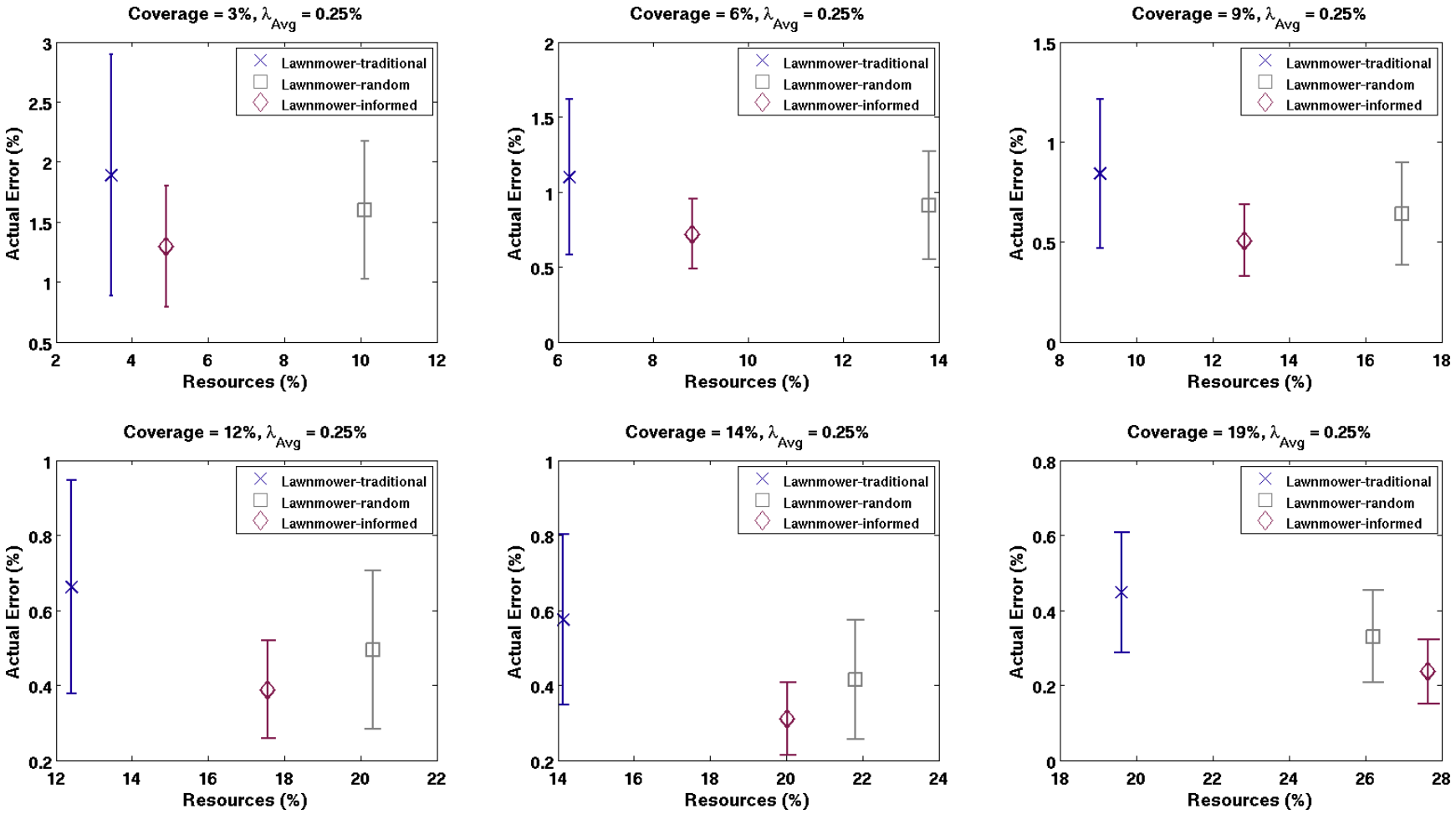


Figure 62: Comparison between navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed, as a function of performance (actual error) and resource (distance) metrics for multiple percent coverages and a specific spatial frequency, λ_{Avg}=0.25%.

Just as observed when comparing these sampling schemes applied to simulated data, lawnmower-informed navigation produces lower error than lawnmower-traditional while requiring less resources than lawnmower-random. We acknowledge the overlap in standard deviation between each algorithm for different coverages, yet, if more data maps are acquired, we project that the distribution will be similar to that of Figure 50. The data supporting the visual comparison between navigation algorithms is provided in Table 9.

Table 9: Relevant performance and resource data for navigation strategies lawnmower-traditional, lawnmower-random, and lawnmower-informed navigation for specific coverages presented in Figure 62.

	Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)	Avg. Resources (%)
Lawnmower-traditional		1.891	1.007	0.713	3.446	4.008	3.446	3.446
Lawnmower-random	3.320	1.598	0.576	0.731	10.036	2.709	10.190	10.190
Lawnmower-informed		1.297	0.503	0.704	4.874	2.334	4.874	4.887
Lawnmower-traditional		1.100	0.516	0.480	6.238	2.143	6.238	6.238
Lawnmower-random	5.975	0.910	0.362	0.375	13.801	1.868	13.801	13.801
Lawnmower-informed		0.718	0.232	0.401	8.822	1.324	8.822	8.828
Lawnmower-traditional		0.841	0.375	0.388	9.054	1.637	9.054	9.054
Lawnmower-random	8.631	0.639	0.257	0.280	16.958	1.332	16.958	16.958
Lawnmower-informed		0.508	0.180	0.247	12.804	1.024	12.804	12.830
Lawnmower-traditional		0.662	0.284	0.326	12.403	1.343	12.403	12.403
Lawnmower-random	11.951	0.495	0.211	0.235	20.308	1.102	20.308	20.308
Lawnmower-informed		0.388	0.131	0.201	17.541	0.687	17.541	17.571
Lawnmower-traditional		0.575	0.228	0.302	14.142	1.108	14.142	14.142
Lawnmower-random	13.942	0.415	0.160	0.206	21.808	0.838	21.808	21.808
Lawnmower-informed		0.311	0.098	0.151	20.000	0.540	20.000	20.036
Lawnmower-traditional		0.448	0.160	0.254	19.612	0.808	19.612	19.612
Lawnmower-random	19.254	0.330	0.125	0.158	26.234	0.586	26.169	26.169
Lawnmower-informed		0.236	0.086	0.103	27.472	0.460	27.672	27.660

Comparison between these algorithms was repeated for the additional set of 10 data maps, representing soil moisture, extracted from MODIS data (Figure 63).

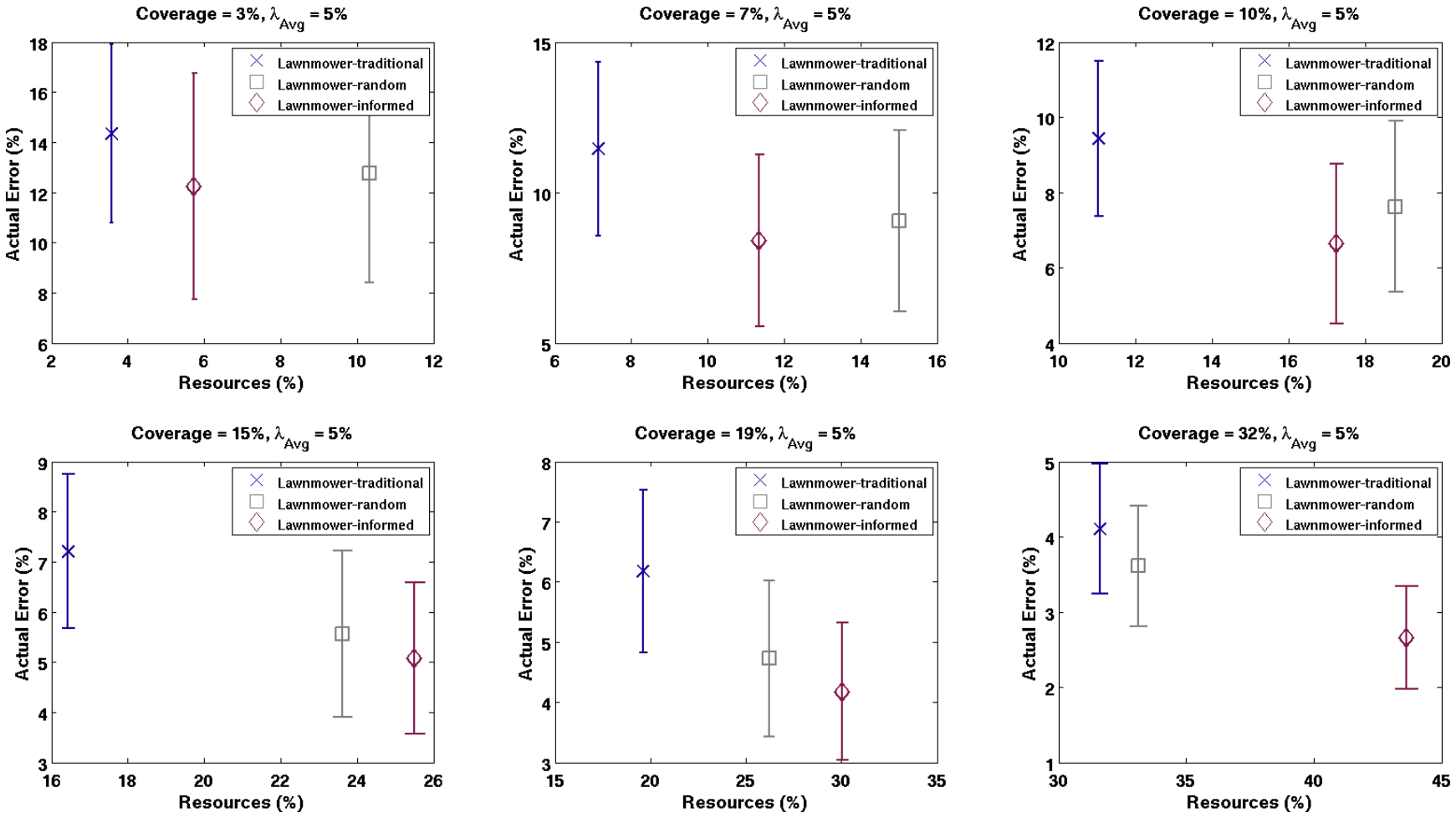


Figure 63: Comparison between navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed, as a function of performance (actual error) and resource (distance) metrics for multiple percent coverages and a specific spatial frequency, λ_{Avg}=5%.

Again, for all coverages presented, the lawnmower-informed method yields an average expected error lower than lawnmower-traditional, while requiring less resources than lawnmower-random. Since a qualitative representation lacks specific numerical detail, Table 10 displays, for select coverages, the quantitative differences between each method with respect to performance (error) and resources (distance).

Table 10: Relevant performance and resource data for navigation strategies lawnmower-traditional, lawnmower-random, and lawnmower-informed navigation for specific coverages presented in Figure 63.

	Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)	Avg. Resources (%)
Lawnmower-traditional		14.350	3.574	10.155	3.569	20.854	3.569	3.569
Lawnmower-random	3.418	12.767	4.366	8.227	10.315	22.064	10.315	10.315
Lawnmower-informed		12.244	4.521	7.887	5.735	21.611	5.761	5.740
Lawnmower-traditional		11.445	2.887	8.083	7.125	17.422	7.125	7.125
Lawnmower-random	6.836	9.054	3.026	6.004	15.020	16.068	15.020	15.020
Lawnmower-informed		8.395	2.865	4.991	10.908	14.662	11.609	11.347
Lawnmower-traditional		9.430	2.052	6.571	11.043	13.202	11.043	11.043
Lawnmower-random	10.253	7.618	2.274	4.915	18.801	11.947	18.801	18.801
Lawnmower-informed		6.635	2.132	4.053	17.282	11.062	16.466	17.252
Lawnmower-traditional		7.206	1.534	4.620	16.440	9.346	16.440	16.440
Lawnmower-random	15.380	5.564	1.664	3.478	23.612	8.696	23.612	23.612
Lawnmower-informed		5.075	1.506	2.858	26.854	7.766	24.981	25.507
Lawnmower-traditional		6.174	1.353	4.038	19.612	8.045	19.612	19.612
Lawnmower-random	18.798	4.729	1.295	2.813	26.201	6.715	26.201	26.201
Lawnmower-informed		4.178	1.142	2.435	29.966	5.966	30.333	30.038
Lawnmower-traditional		4.101	0.864	2.578	31.623	5.006	31.623	31.623
Lawnmower-random	31.614	3.614	0.801	2.329	33.128	4.620	33.128	33.128
Lawnmower-informed		2.664	0.683	1.474	42.714	3.546	44.494	43.629

The data reflects a dominant preference towards the use of lawnmower-informed navigation over lawnmower-traditional and lawnmower-random. This preference is based on its average performance of low expected error and low average resources relative to the other navigation options. This observation is consistently made for coverages less than 15 percent for the DEMs tested with average spatial frequencies equal to 0.25 percent and coverages less than 10 percent for the data maps tested whose average spatial frequencies are equal to 5 percent. Still, depending on the user's requirements for maximum expected error ($error_{Max}$) and maximum resources (D^{NN}_{Max}), lawnmower-traditional could be considered the preferred navigation option, especially if the resource weighting factor, α_R, is much greater than the performance weighting factor, α_P (Section 3.4).

5.3.1.2 Performance of Informed Navigation on Realistic DEMs for Continuous Spaces

As was presented in Section 5.2.2.1, we include the application the trajectories of each navigation method, recorded from navigation in a quantized space, to our set of 20 DEMs (in their original continuous domain). After extracting these locations, applying our quadratic interpolation method, and calculating the error between the resulting reconstructed DEMs and the ground truth, we record the corresponding RMS error for each sampling configuration (Figure 64).

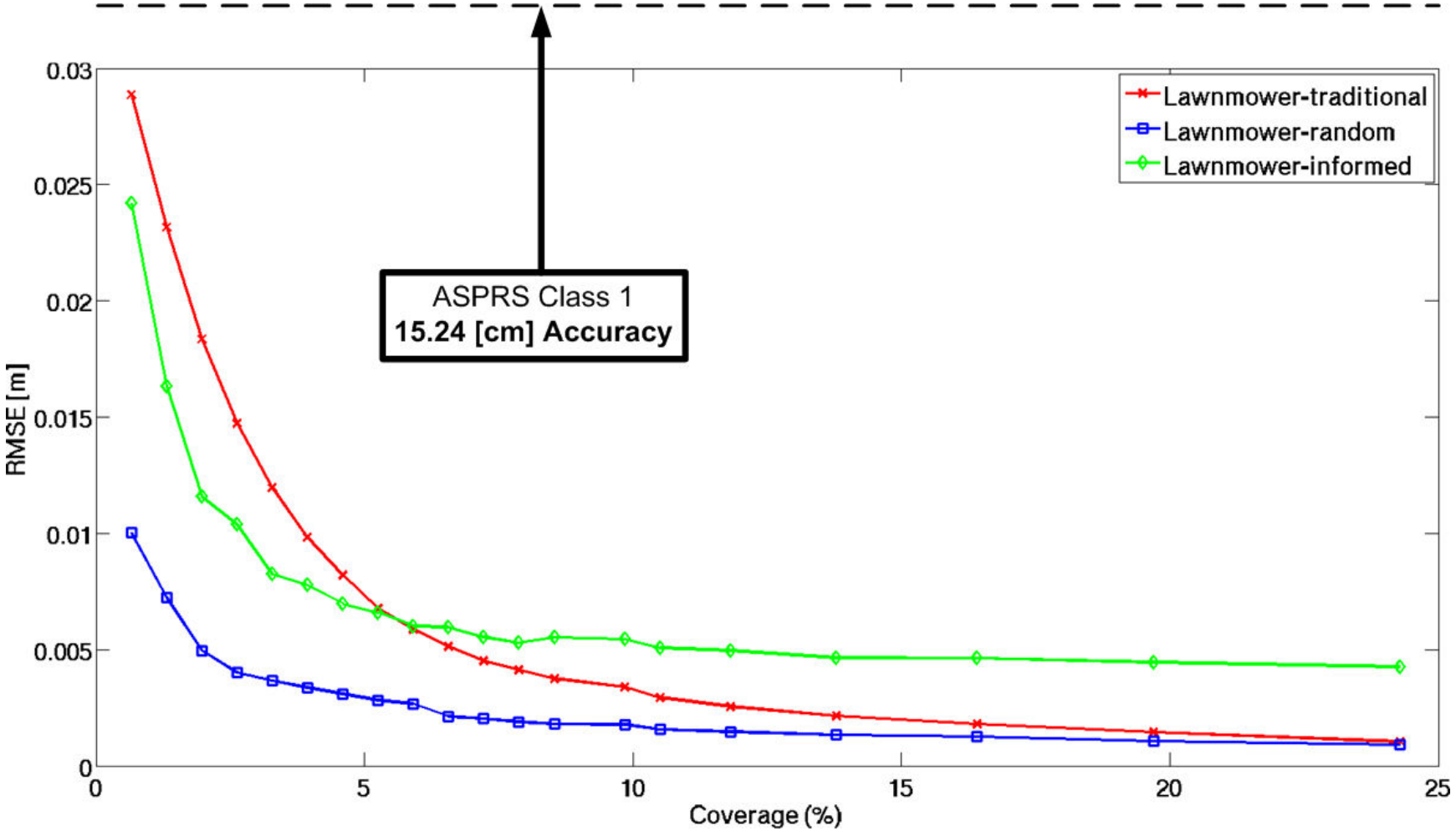


Figure 64: Comparison of RMS errors, produced by navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed to the ASPRS Class 1 map accuracy standard, as a function of coverage in a continuous domain (realistic DEMs).

For this analysis, at all coverages, regardless of the navigation algorithm chosen to select sampling configurations, the resulting RMS error meets the highest ASPRS mapping standard (Class 1 map accuracy). A similar range of performance improvement by lawnmower-informed over lawnmower-traditional exists (between 0 and 5 percent) as was shown for simulated DEMs (Figure 54), however, a scientist need only consider that improvement if finer error requirements need to be met. This improvement of lawnmower-informed over lawnmower-traditional, again, validates the benefit of leveraging sampled data during navigation to inform subsequent sampling decisions within the sample space. Since ASPRS map standards are not valid for soil moisture data, we will not repeat this validation for the set of MODIS data.

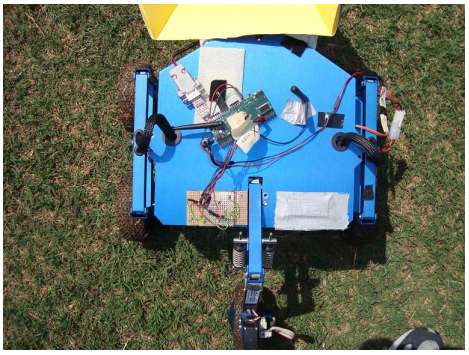
We conclude this chapter by applying our estimation methodology and lawnmower-based navigation sampling schemes to a terrain modeled with data collected during a robotic survey of a real space.

5.3.2 Applied Metrics to Real Environments

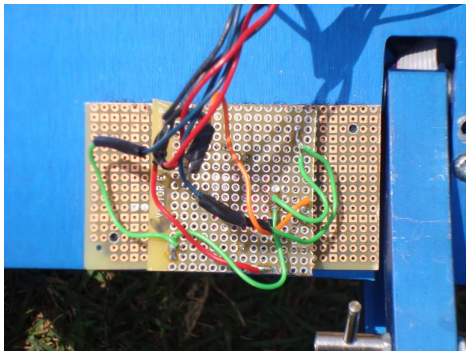
As a final analysis, we applied sampling configurations generated by our three lawnmower-based navigation algorithms on a terrain modeled from data collected in the physical world. Using a wheeled robotic platform capable of traversing variable terrain, we mounted equipment to perform the necessary data collection within our test environment. The SECT-II platform, commercially available from Bluebotics (`www.bluebotics.com`), was employed to perform a topographical survey. Designed primarily for dry terrain, this six-wheeled system supports navigation over a variety of spatially complex surfaces ranging from slowly varying hills to jagged, rocky surfaces, *i.e.*, multiple spatial frequencies. We outfitted the SECT-II with a low-cost dual-axis micro-electrical mechanical systems (MEMS) accelerometer from Analog Devices (ADXL322) capable of measuring at angular resolutions of 8.31 [mV/degree]

and 8.38 [mV/degree] along its x and y axes, respectively. We also included a commercially available microprocessor to the SECT-II. The Connex 400XM from Gumstix (`www.gumstix.com`) includes a 400MHz ARM processor, wireless 802.11g ethernet, and bluetooth capabilities. We interfaced a Robostix board, also from Gumstix, with the Connex processor, which included an Atmel ATMega 128 RISC microcontroller. The Robostix provided an analog-to-digital converter (ADC) unit that simplified the conversion of continuous analog voltage output from the accelerometer into relevant tilt measurements (Figures 65a-65d). The SECT-II passively collected tilt informa-

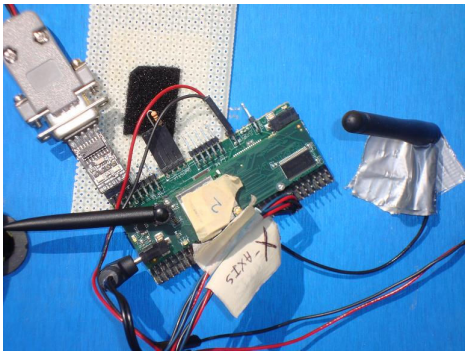
(a) SECT-II robotic platform from Botics Inc.



(b) Top view of SECT-II equipped with sensing equipment.



(c) ADXL accelerometer from Analog Devices Inc.



(d) Gumstix/Robostix μP.

Figure 65: Hardware components used for the robotic survey system.

tion while navigating repeatedly across a 20 [m] x 40 [m] testing area in a local park in Atlanta, GA (Figure 66).

Just as the range of coverages varied with size and resolution for simulated and

(a) Perspective 1. (b) Perspective 2.

Figure 66: Test site for hardware demonstration of the robotic survey system.

realistic data sets, as discussed in Sections 5.2.1.1 and 5.3.1, respectively, this range changed again for our test site and is displayed in Table 11.

Using the continuous pitch and roll values from our robotic survey, we are able to extract a continuous and quantized terrain model on which we can test our navigation schemes, lawnmower-traditional, lawnmower-random, and lawnmower-informed (Figure 67).

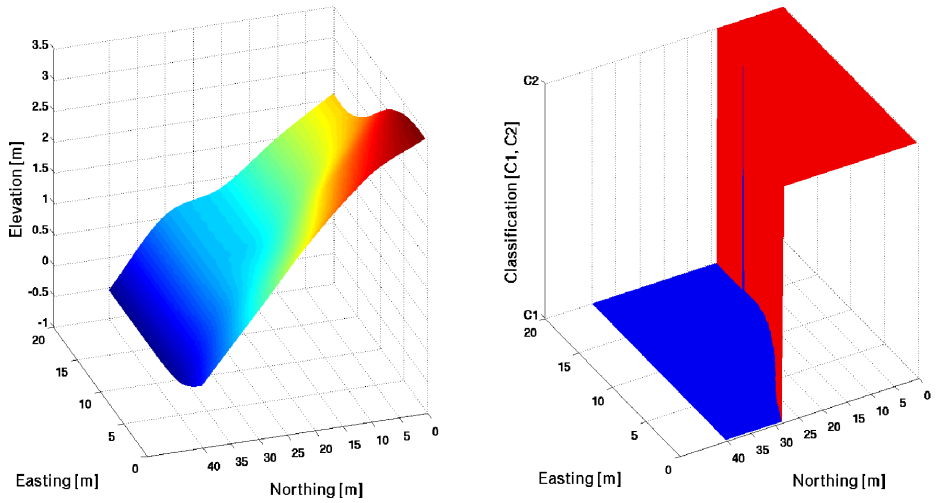


Figure 67: Example of a DEM extracted from robotic survey conducted in the physical world (left) and its quantized version (right).

We look at the performance of each navigation algorithm as a function of multiple configurations at different coverage values when applied to this single DEM. In the final test of the effectiveness of our informed estimation methodology, we apply the

Table 11: Coverage ranges for 20 [m] x 40 [m] test area at centimeter resolution.

Number of Swaths	Sample Quantity, B	Coverage (%)
1	100	0.498
2	200	0.969
3	300	1.434
4	400	1.990
5	500	2.488
6	600	2.985
7	700	3.483
8	800	3.980
9	900	4.478
10	1000	4.975
11	1100	5.473
12	1200	5.970
13	1300	6.468
14	1400	6.965
15	1500	7.463
16	1600	7.960
18	1800	8.955
19	1900	9.453
22	2200	10.945
24	2400	11.940
28	2800	13.930
33	3300	16.418
39	3900	19.403
49	4900	24.378

lawnmower-traditional navigation policy as a sampling scheme (Figure 68).

For a single terrain example, the performance trends of the actual error generated by our informed estimation method outperform that of traditional estimation by more than 40 percent across a majority of the coverage range tested, approximately 1-20 percent. As seen throughout this chapter, because of the low spatial frequency of the DEM tested (λ_{Avg}=1%), there exists a large gap between the expected and actual error generated by our informed estimation method. Again, this allows us to label the expected error as an upper limit on actual error achieved with our methodology for environments with these low spatial frequencies ($\lambda_{Avg} \leq 1.5\%$). We repeat this

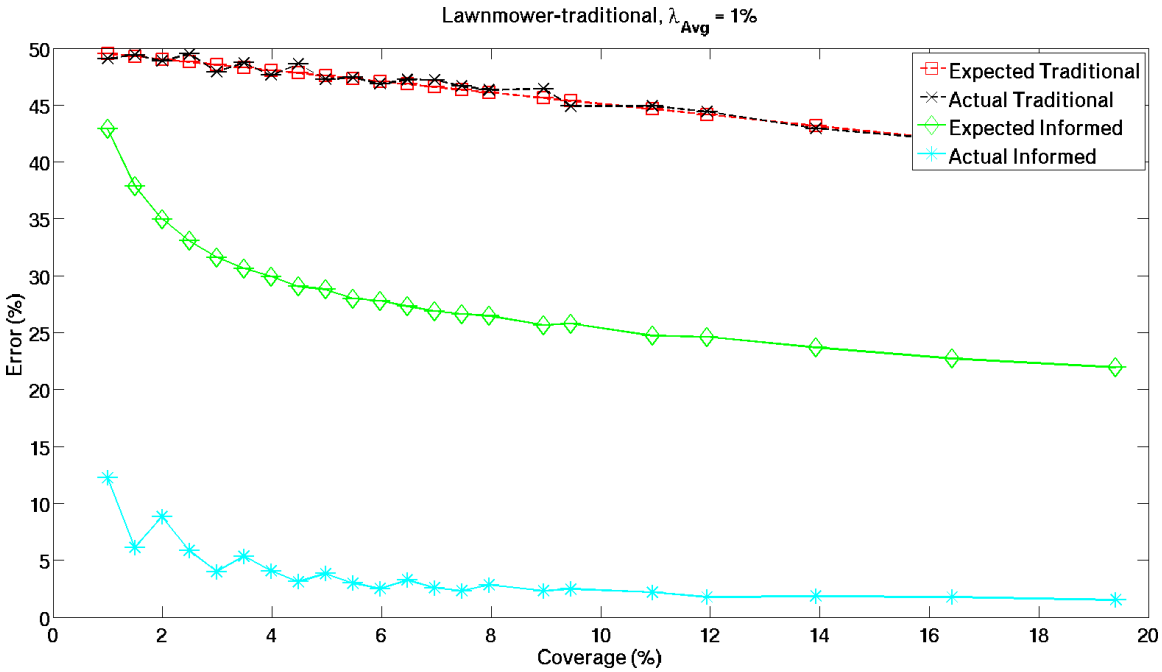


Figure 68: Performance of the lawnmower-traditional navigation strategy across multiple coverages as applied to a DEM extracted from a robotic survey of a real terrain where λ_{Avg}=1%.

test for lawnmower-random navigation (Figure 69).

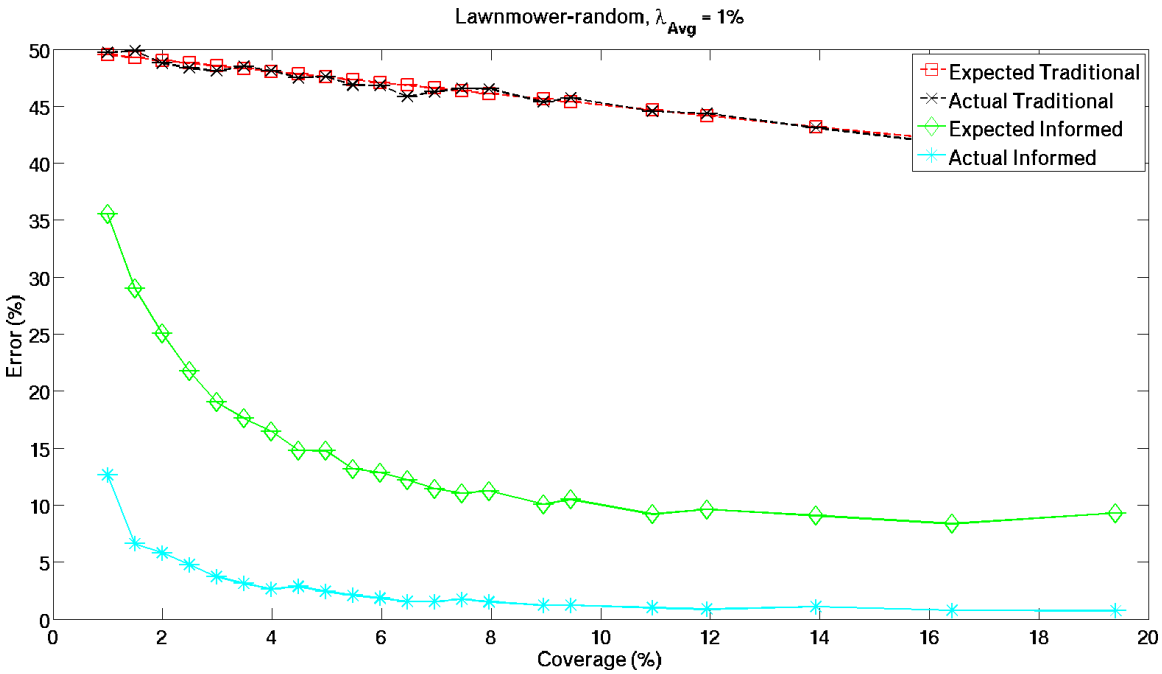


Figure 69: Performance of the lawnmower-random navigation strategy across multiple coverages as applied to a DEM extracted from a robotic survey of a real terrain where λ_{Avg}=1%.

For coverages greater than 8 percent, the gap between expected error and actual error narrows significantly (Figure 69). We attribute this primarily to the improved spatial diversity generated from sampling by lawnmower-random navigation, better enabling our estimation method to more accurately project the classification of unsampled locations.

We conclude the verification of our informed estimation method across a range of coverages, displaying the performance trends associated with samples generated from the lawnmower-informed navigation strategy (Figure 70).

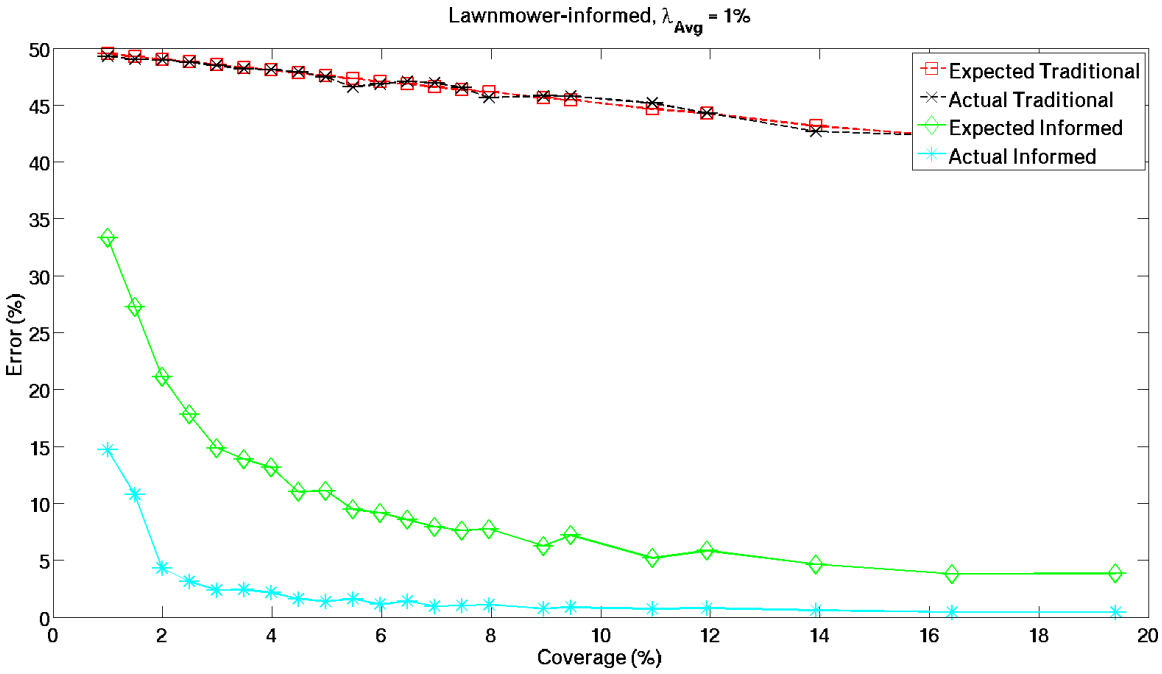


Figure 70: Performance of the lawnmower-informed navigation strategy across multiple coverages as applied to a DEM extracted from a robotic survey of a real terrain where λ_{Avg}=1%.

Just as seen with the results of lawnmower-random, the gap between expected and actual error closes as coverage increases (Figure 70). This gap represents approximately a 5 percent difference whereas the gap generated by lawnmower-random represents at least an 8 percent difference for coverages greater than 11 percent.

For all three performance plots (Figures 68, 69, and 70), the expected and actual error trends of the traditional estimation gradually decrease as coverage increases, but

not noticeably as much as for our informed estimation method. Additional commentary on this difference between the two estimation methods is included in Appendix A, where performance is represented as a function, not of coverage, but of percentage configuration estimation confidence, η_{Avg}.

5.3.2.1 Comparison between navigation algorithms: Constraint analysis

We contrast the performance of each navigation algorithm as a function of its required resources as applied to our real terrain example in the quantized domain (Figure 71).

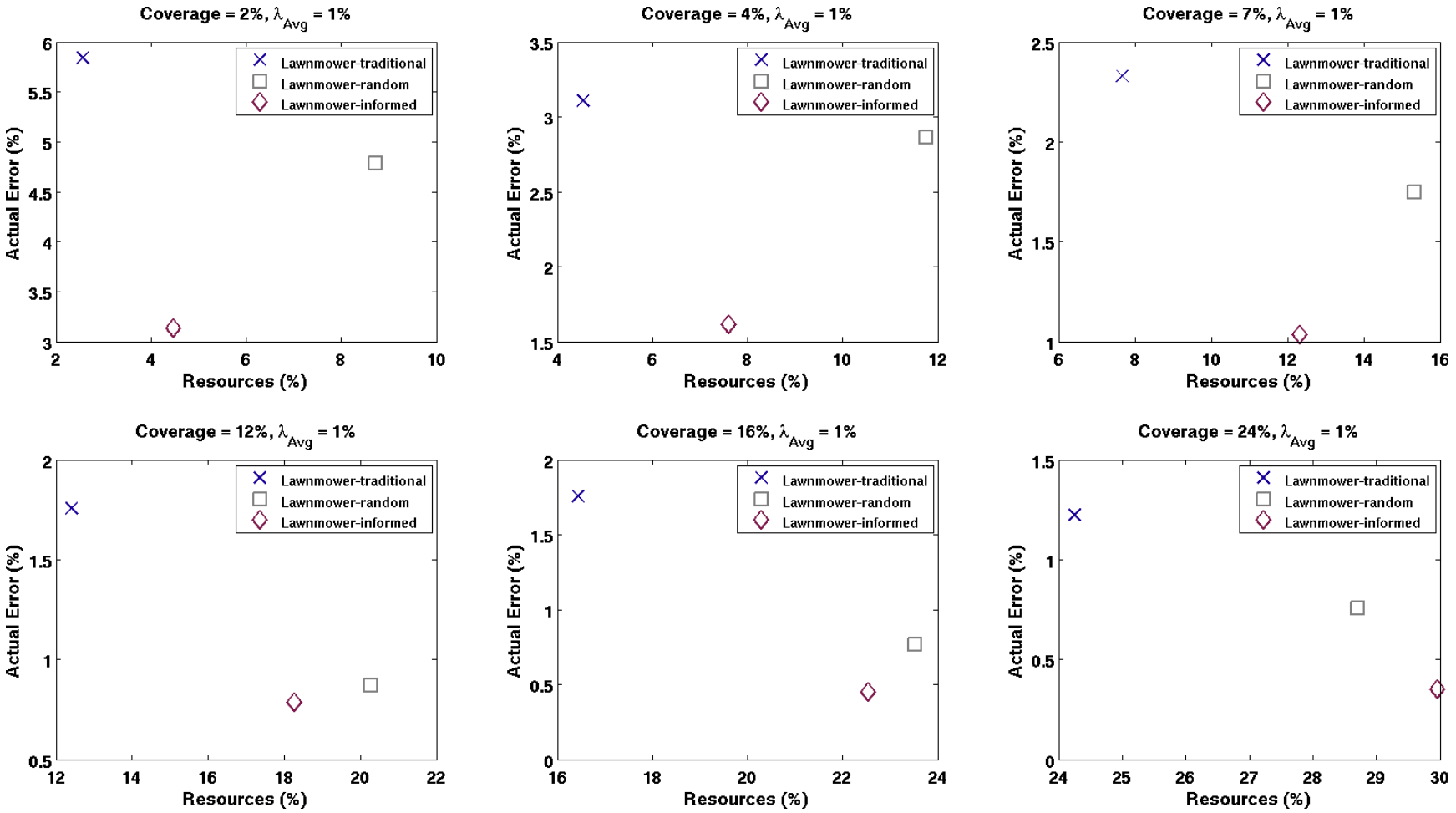


Figure 71: Comparison between navigation algorithms, lawmower-traditional, lawnmower-random, and lawnmower-informed, as a function of performance (actual error) and resource (distance) metrics for multiple percent coverages and a specific spatial frequency, λ_{Avg}=1%.

For this set of results, only a single terrain was available for testing. Yet, in assessing the performance of these sampling schemes, lawnmower-informed navigation outperforms the lawnmower-traditional and lawnmower-random methods with respect to actual error and distance, respectively. Data for the navigation algorithms as applied to the single terrain is shown in Table 12.

Table 12: Relevant performance and resource data for navigation strategies lawnmower-traditional, lawnmower-random, and lawnmower-informed navigation for specific coverages presented in Figure 71.

	Coverage (%)	Error (%)	Resources (%)
Lawnmower-traditional		5.841	2.563
Lawnmower-random	2.438	4.786	8.721
Lawnmower-informed		3.134	4.477
Lawnmower-traditional		3.109	4.541
Lawnmower-random	4.388	2.866	11.752
Lawnmower-informed		1.612	7.609
Lawnmower-traditional		2.328	7.670
Lawnmower-random	7.313	1.746	15.328
Lawnmower-informed		1.035	12.341
Lawnmower-traditional		1.756	12.403
Lawnmower-random	11.701	0.871	20.268
Lawnmower-informed		0.786	18.263
Lawnmower-traditional		1.756	16.440
Lawnmower-random	16.090	0.766	23.529
Lawnmower-informed		0.453	22.540
Lawnmower-traditional		1.224	24.254
Lawnmower-random	23.891	0.756	28.704
Lawnmower-informed		0.353	29.961

As an example application of how to identify the navigation option of choice, we first switch perspectives from the quantized to continuous domain. As was presented in Sections (5.2.2.1 and 5.3.1.2), we will apply the trajectories originally generated by lawnmower-traditional, lawnmower-random, and lawnmower-informed from a quantized terrain and apply them to the continuous terrain map, recording continuous values at each (x, y) sample location of each configuration. After applying quadratic 3D

interpolation to the sample configurations associated with each navigation method, we can calculate a measure of RMS error for each method and use an ASPRS mapping standard (Section 5.1.2) as our performance constraint. We use ASPRS map accuracy standards in lieu of defining an arbitrary value for $error_{Max}$ (Section 3.4).

For our 20 [m] x 40 [m] space, at a between-sample resolution, δ, of 0.2 [m], we assume the sampling operation allows for 2400 samples (or 12 percent coverage) to be collected during navigation. Given the specifications of our SECT-II platform, we define its ideal operating velocity as 0.5 [m/s]. We define a specific amount of battery time (25 minutes) and convert to the equivalent amount of resources necessary for achieving 12 percent coverage at 0.5 [m/s], defined as D_{Max}^{NN} (Equation 31).

$$D_{Max}^{NN} = \frac{(\text{Velocity [m/s]})(\text{Battery Time [s]})}{(\text{Number of Samples})} \; [m] \tag{29}$$

$$= \frac{(0.5)((25)(60))}{(2400)} \; [m] \tag{30}$$

$$= 0.3125 \; [m] \tag{31}$$

We determine the analogous swath bandwidth required for 2400 samples, D_{BW}^{NN}, as approximately 1.6 [m], which provides our resource constraint of 19.53 percent using Equation (31). We summarize these two constraints below (Table 13). Next,

Table 13: User-defined performance and resource constraints based on ASPRS Class 1 map accuracy standard and D_{Max}^{NN}, respectively.

User-defined Constraint	Definition	Value
1. Performance	ASPRS Class 1 Accuracy	15.24 [cm]
2. Resources	$100(D_{Max}^{NN}/D_{BW}^{NN})$	19.53%

we identify where each navigation algorithm succeeds or fails in meeting these user constraints (Figure 72).

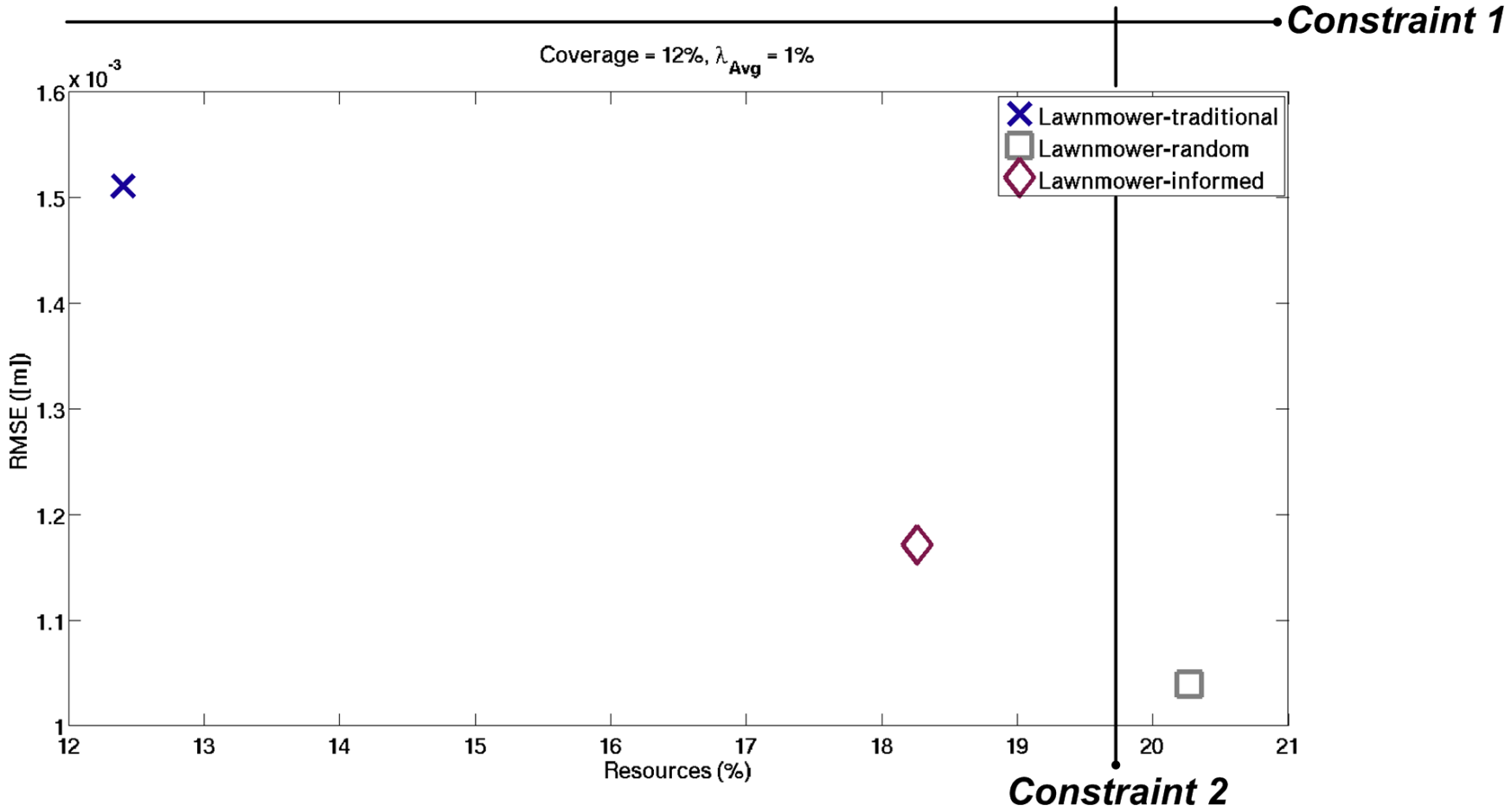


Figure 72: Application of user-defined constraints from Table 13 to constraint analysis plot contrasting lawnmower-traditional, lawnmower-random, and lawnmower-informed navigation algorithms as sampling strategies.

In the event that a quantitative ranking is required, we can defer to Equation (21) and use the value calculated for ψ, should more algorithms need to be compared against one another.

Continuing along the same lines of evaluating these navigation strategies in the continuous domain, we present the performance of each strategy for a range of coverages (Figure 73).

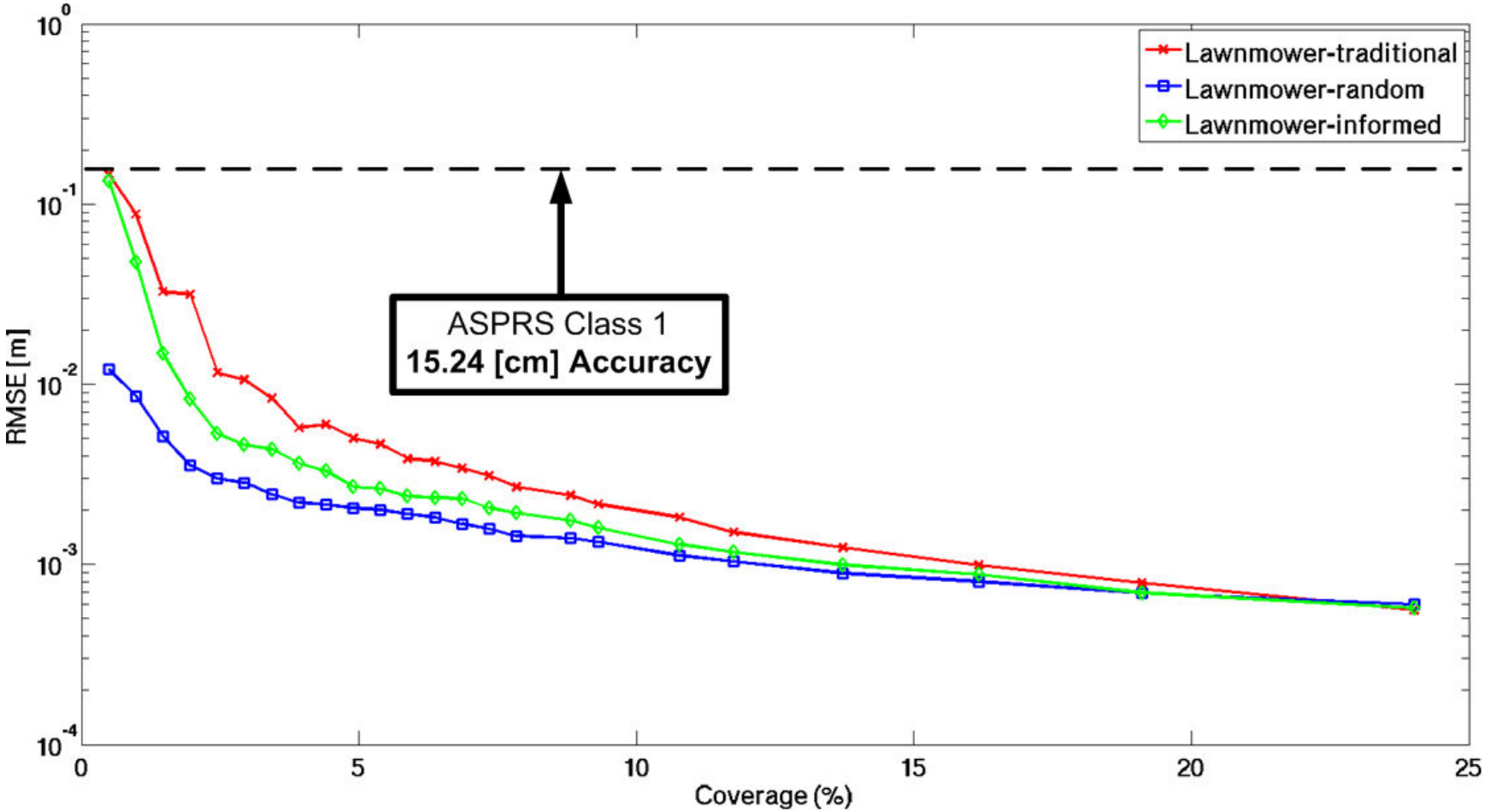


Figure 73: Performance of navigation algorithms, lawnmower-traditional, lawnmower-random, and lawnmower-informed across a range of coverages, meeting the ASPRS Class 1 map accuracy standard as an error constraint.

Consistent with the RMS error presented at a single coverage in Figure 72, across a larger range of coverages, the lawnmower-informed navigation, used as a sampling strategy, generates reconstruction error less than that of lawnmower-traditional. Regardless, for all coverages considered, each navigation option is capable of producing sampling configurations that result in reconstruction error meeting ASPRS Class 1 map accuracy standards.

CHAPTER VI

CONCLUSION

6.1 Concluding Remarks

Our research addresses issues related to improving Earth-observing systems through robotics for the purpose of providing better geodetic information. By considering the intersecting areas of spatial sampling and estimation, we have integrated considerations relevant to a robotic survey system. Specifically, our work provides an Earth scientist with the following contributions: 1) An improved method for estimating spatially distributed phenomena, 2) the theory and application of a navigation algorithm as a useful sampling strategy (informed by the estimation method) that shows improvement over other baseline navigation strategies used to collect samples, and 3) a method for comparing these different navigation algorithms based on their resulting performance (*i.e.*, their ability to generate sampling configurations that produce low reconstruction error) and required resources.

These contributions are increasingly relevant to garnering a better understanding of geo-physical changes taking place within different environments. Our work is especially applicable to environments whose information is characterized as belonging to one of two classes, *i.e.*, a quantized space, but can also be applied to continuous spaces. By leveraging the benefits afforded by this research in environmental robotics, advances in areas of sampling and estimation can take place more readily for future applications. Equipped with relevant robotic platforms, capable of sampling the area of interest, scientists will still need to quantify the benefit of different sampling schemes executed by these robotic systems. This research enables the process of quantifying which sampling configurations to consider and why based on realistic

constraints.

6.2 Recommendations for Future Work

Throughout our work, we presume to have perfect knowledge for our experiments. This perfect knowledge includes accurate readings from the sensor used to collect measurements at each of the B locations in S. We also condition our work on the perfect knowledge of the mobile agent's global position in S, and accurately arrive at the correct location of each subsequent sampling location during navigation. Using this perfect knowledge that we have set forth is still a valid approach in theory and in simulation, yet, inaccurate (or noisy) sensor readings, and error-prone positioning systems, must be accounted for during implementation of a field-ready robotic survey system. Considering these types of real errors is true both during intelligent navigation and offline estimation. We learn from Moghaddam *et al.* of the challenge in properly modeling noisy measurements of specific sensors and the resulting impact on the accuracy of readings taken during a field experiment [69]. Based on the reality of faulty sensors, their characterization should be considered during the sampling and estimation process. Williams *et al.* discuss the value of accurate localization methods for a mobile sensor platform, especially when navigating in terrains with harsh weather conditions [15]. Any deployable robotic survey system should incorporate position information as feedback to the agent throughout the sampling task to ensure that the measurements taken are properly assigned to a correct (x, y) location within the pixel footprint. Pursuant to the issues of sensor and localization error, based upon the theory in Chapters 3 and 4 and the results presented in Chapter 5, there are several improvements that will serve to greatly expand each of the contributions in this work.

With respect to the first contribution, we propose additional work for the informed estimation method described. In Sections 5.2.1.1-5.3.2, we confirmed a significant

improvement in actual error when estimating a space using our informed estimation method versus that of traditional, single-trial Bernoulli estimation (between 30-40 percent for a specific range of coverage). An improvement to this informed estimation method, however, includes reducing the difference between the expected and actual error generated for very low spatial frequencies. We state, in Chapter 5, that for terrain models of low spatial frequency ($\lambda_{Avg} \leq 1.5\%$), the expected error generated by our informed method serves as an upper limit on actual error. At higher spatial frequencies ($\lambda_{Avg} = 25\%$), we actually see a closer approximation of actual error by the expected error curve (Appendix B). The robustness of our method will be significantly improved if the approximation of actual error by expected error was consistent for a terrain model of any spatial frequency. Next steps require incorporating the value of λ_{Avg} into the estimation theory (Chapter 3). It is believed that completing this step will improve the estimation method's ability to accurately estimate actual error.

The second recommended improvement to our work relates to the actual navigation algorithms chosen as sampling schemes. We proved as a part of our third contribution, that, with lawnmower-informed navigation, by using our informed estimation theory to make more useful sampling decisions, the resulting average reconstruction error is lower than that of lawnmower-traditional navigation. Concurrently, if lawnmower-informed navigation conservatively sacrifices an increase in distance traversed to obtain better samples, the expended resources are also less than that required to sample randomly according to lawnmower-random navigation. A potential improvement to the lawnmower-informed navigation, however, would be to incorporate a heuristic associated with the data. Similar to our first recommendation to incorporate spatial frequency into the estimation theory, using a heuristic to guide navigation has shown promise in similar work contrasting performance and resource constraints [56].

We recall the set of spatially-distributed trajectories with an underlying lawnmower structure from Bourgeois's work in [7] (Section 2.4). When we use the regularly-spaced linear swaths as a reference around which an agent can navigate according to a heuristic, a more informed coverage-driven navigation policy is generated. By collecting the local information within the neighborhood of each swath of a symmetrically placed lawnmower pattern, the potential of better estimating the entire sample space is increased [56]. Based on these recommendations, there exist multiple opportunities to improve knowledge about the relationship between navigation, sampling, and estimation.

The final recommendation for this work relates to the constraint analysis. As our third contribution, we provide analysis of the performance and required resources of three distinct navigation algorithms used to generate unique sampling configurations. The two constraints of performance and resources, however, need not be the only criteria by which a navigation algorithm is evaluated. The number of constraints can easily be increased from a two-dimensional to n-dimensional evaluation space, depending on the application or phenomena under test.

This work has provided the theoretical framework for an informed estimation process, an informed navigation policy driven by that estimation process, and constraint-based comparisons between sampling schemes. Our additional perspective allows the robotics and geo-science communities to collaboratively design more relevant technologies that improve information gain about geo-physical phenomena. With the advent of these technologies, more useful geodetic information for the future will be gained, in more locations, and at greater spatial resolutions within the pixel footprint.

APPENDIX A

PERFORMANCE VS. η_{AVG}

A.1 Simulated Terrain Models

Here, we include all supplemental content related to plotting error as a function of configuration confidence (η_{Avg}) for environmental phenomena exhibiting low spatial frequencies (λ_{Avg}=1.5%). When the number of samples per configuration is held constant, for each terrain to which a specific navigation algorithm is applied, the sampling configuration generated produces its own unique estimation confidence value, η_{Avg}. When we sort and rank each configuration, then look at the corresponding performance (expected and actual) we continue to observe a similar trend of our informed estimation method outperforming the single-trial Bernoulli (or traditional) estimation method. Plotting performance as a function of η_{Avg} compensates for the inability to compare configurations as a function of coverage since the number of samples in each configuration is constant. We expand on the results of Section 5.2 by evaluating the error produced by each unique configuration generated at a constant amount of coverage, *i.e.*, each configuration containing B samples. We have shown in previous sections, that even when the number of samples is set as constant, there exist variability between configurations as it relates to the amount of expected error that is generated. This variability in error is attributed to both the unique spatial placement of each sample in a given configuration and that sample's classification as it relates to the estimation of an unknown location. We set coverage equal to approximately 30 percent of the total number of possible samples, per our earlier observations (Section 3.2.2), and plot error as a function of a configuration's percent estimation confidence, $100(\eta_{Avg})$ (Equation (19)). Again, we are interested in the relationship between the

expected error projected by our informed estimation method and the actual error generated based on our informed estimation rule (Section 3.3). We include the error (expected and actual) generated by the traditional estimation as a reference and present data for all navigation algorithms used as sampling schemes, keeping λ_{Avg} constant for all terrains tested. The set of Q_B configurations tested here are associated with B samples that constitute approximately 32 percent coverage as presented in Table 6 (Figure 74).

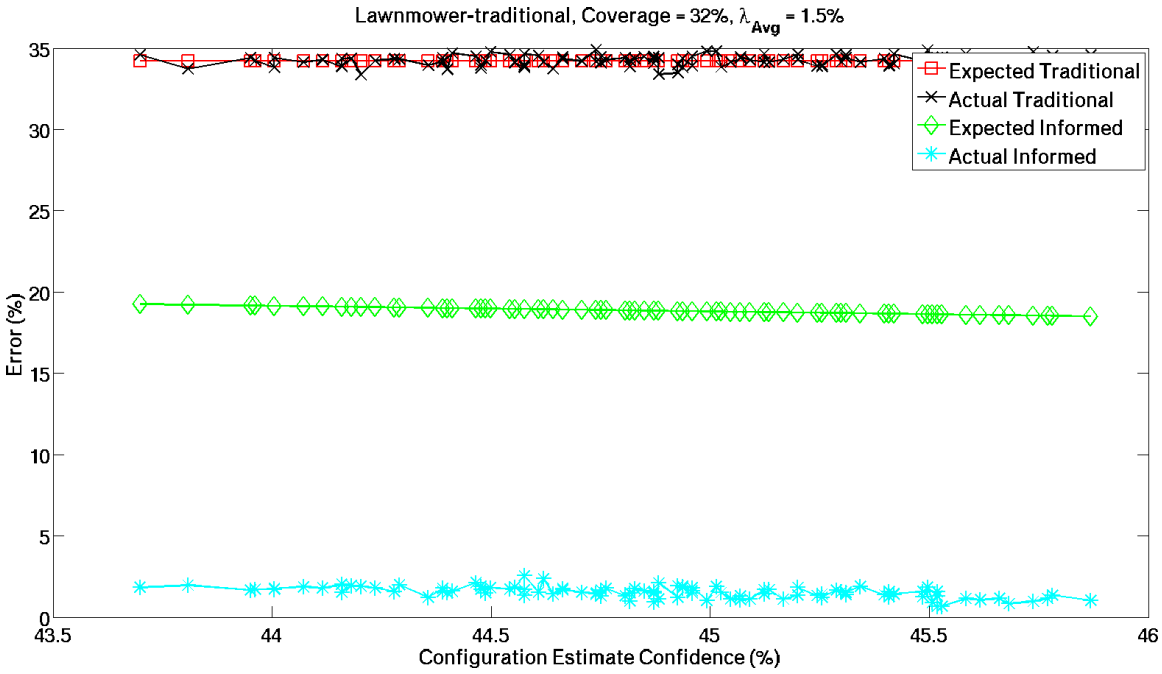


Figure 74: Error (expected and actual) versus percentage of configuration estimate confidence, η_{Avg}, for lawnmower-traditional navigation algorithm at 32% coverage and λ_{Avg}=1.5%.

For the sample configurations generated by lawnmower-traditional, the gap between expected and actual error (as produced by our informed estimation method) is more than 15 percent for DEMs with a spatial frequency of 1.5 percent.

While the performance trends are useful, the more relevant observation from these plots is related to our configuration estimation confidence metric, η_{Avg}. In Chapter 3, we generated analogous results, comparing expected and actual error as a function

of the raw confidence value, η_{Avg}, for smaller spaces, $MN = 9$ and $MN = 16$ (Figures 18-20). Here, in Figure 74, we plot error values as a function of a confidence percentage. This percentage is calulated by dividing the raw configuration estimation confidence of a particular configuration by the maximum confidence possible. This maximum confidence is equal to 1 multiplied by the quantity of all estimated locations in the space, $MN - B$. Performance as a function of η_{Avg} is generated for the configurations produced by lawnmower-random navigation with resulting trends shown below (Figure 75).

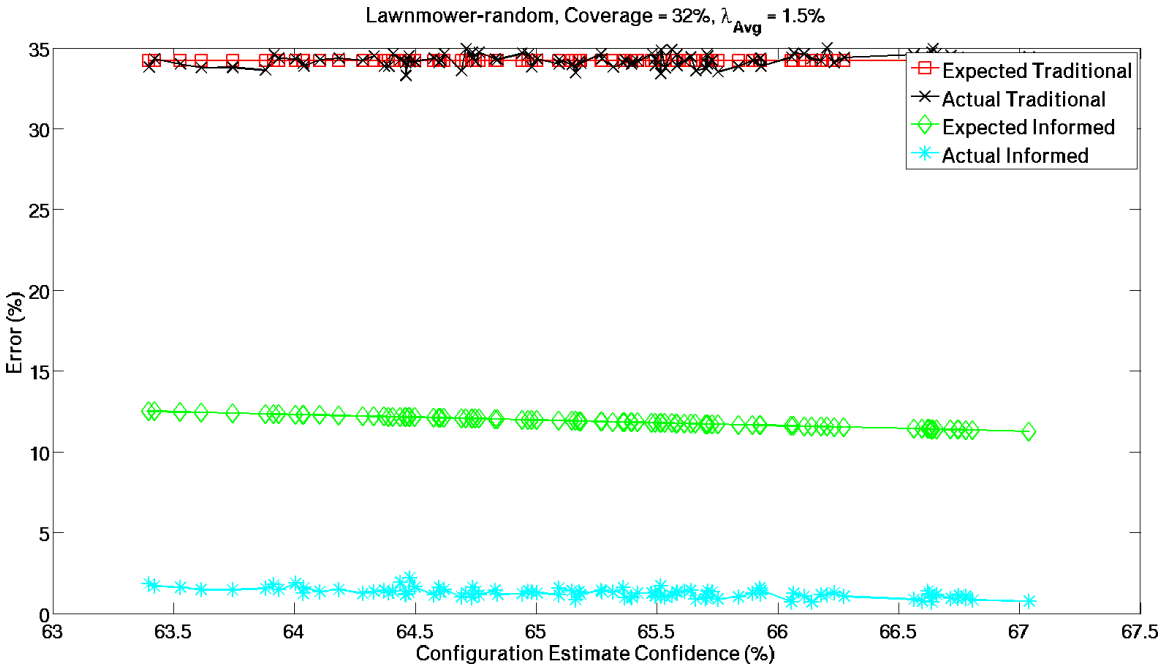


Figure 75: Error (expected and actual) versus percentage of configuration estimate confidence, η_{Avg}, for lawnmower-random navigation algorithm at 32% coverage and λ_{Avg}=1.5%.

Again, as seen with lawnmower-traditional, the actual error generated by our informed estimation method falls below the expected error for lawnmower-random related configurations and outperforms traditional estimation. We mention this consistency in performance between unique configurations to emphasize that the trends associated with our informed estimation method are the same whether all Q_B configurations are tested or only a smaller subset. For these terrains with low spatial

frequency ($\lambda_{Avg} = 1.5\%$), a 10 percent gap exists between expected and actual error trends (Figure 75). We apply one more set of configurations associated with lawnmower-informed navigation to our test DEMs (Figure 76).

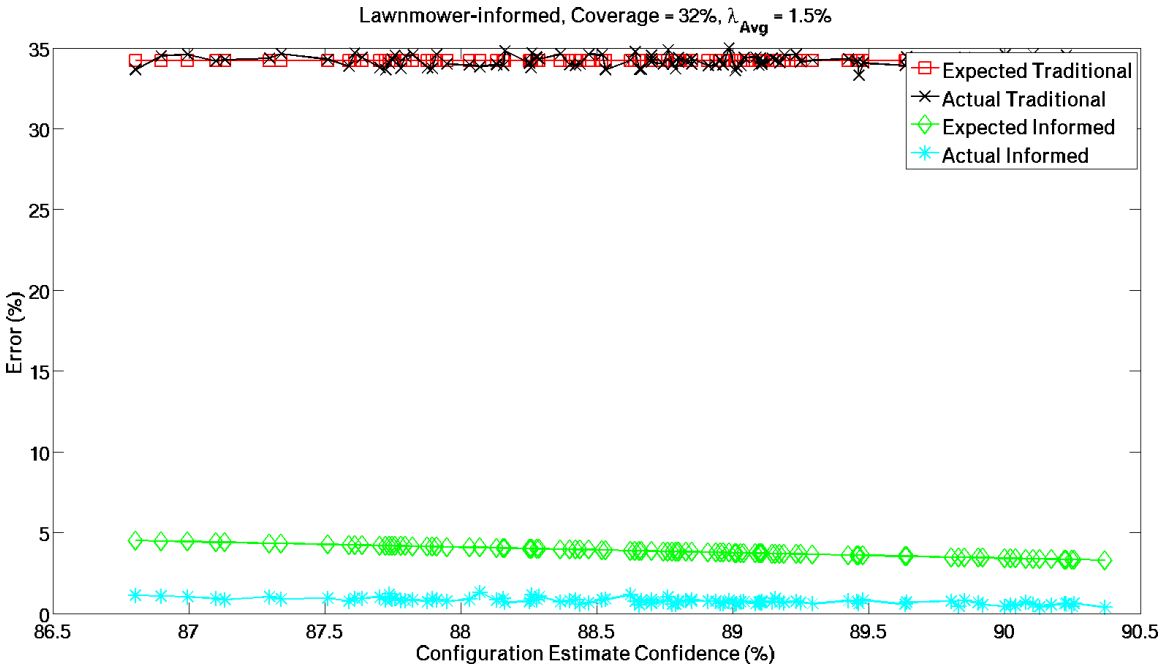


Figure 76: Error (expected and actual) versus percentage of configuration estimate confidence, η_{Avg}, for lawnmower-informed navigation algorithm at 32% coverage and λ_{Avg}=1.5%.

As with the trends in Figures 74 and 75, at the smallest spatial frequency (1.5 percent), our informed estimation method approximates an upper limit on actual error with a 5 percent gap between trends.

APPENDIX B

IMPACT OF λ_{AVG} ON PERFORMANCE

This appendix includes all data related to the impact of data maps exhibiting moderate and high spatial frequencies on the performance of our informed estimation method. For terrains whose spatial frequency is equal to or greater than 25 percent, the performance of our estimation method begins to change and the expected error generated no longer provides an upper limit on the actual error. This change is observed for each of the navigation methods used to generate unique sampling configurations for these test spaces (Figures 77-82).

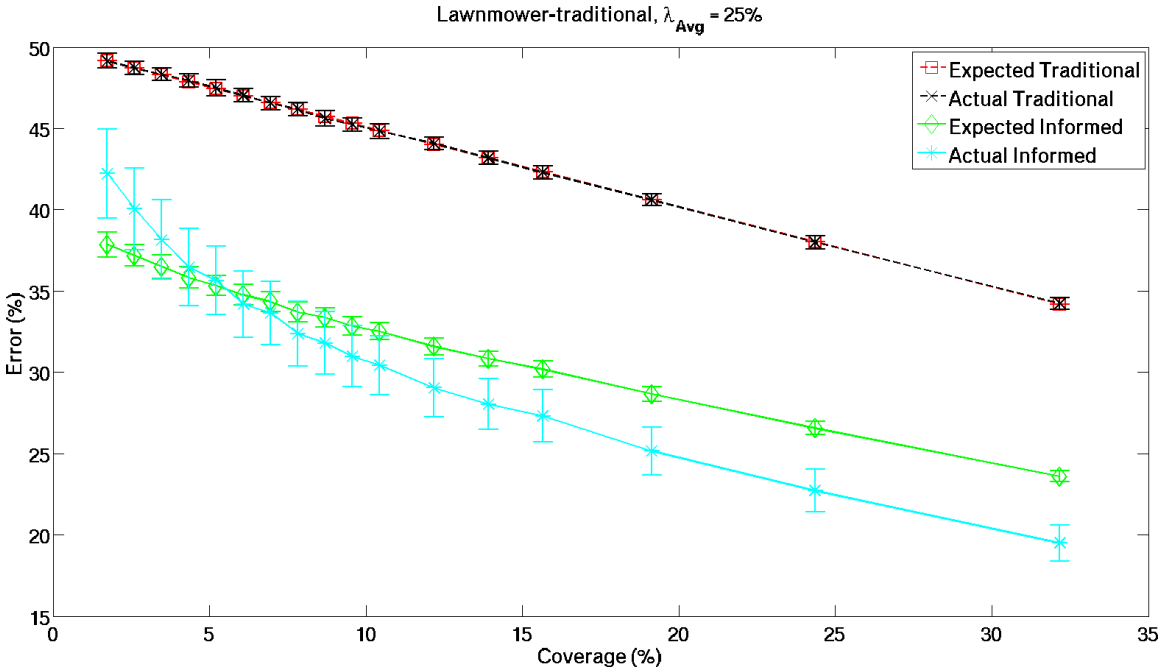


Figure 77: Performance of informed estimation using lawnmower-traditional navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=25%.

As the spatial frequency is increased to 25 percent, our informed estimation

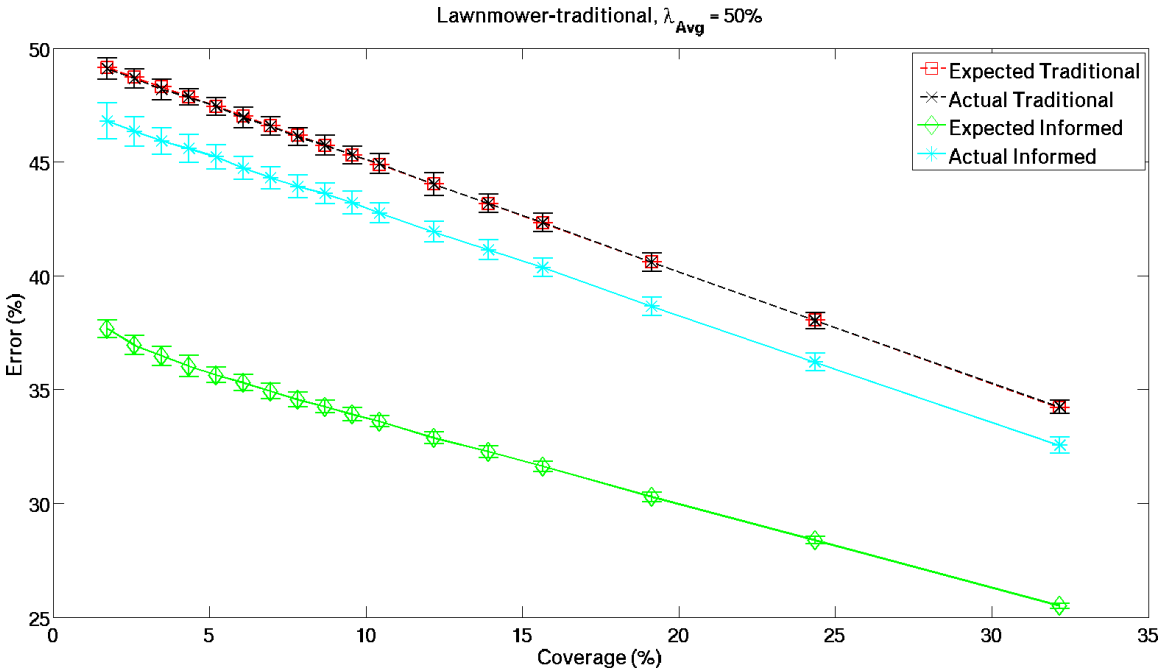


Figure 78: Performance of informed estimation using lawnmower-traditional navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=50%.

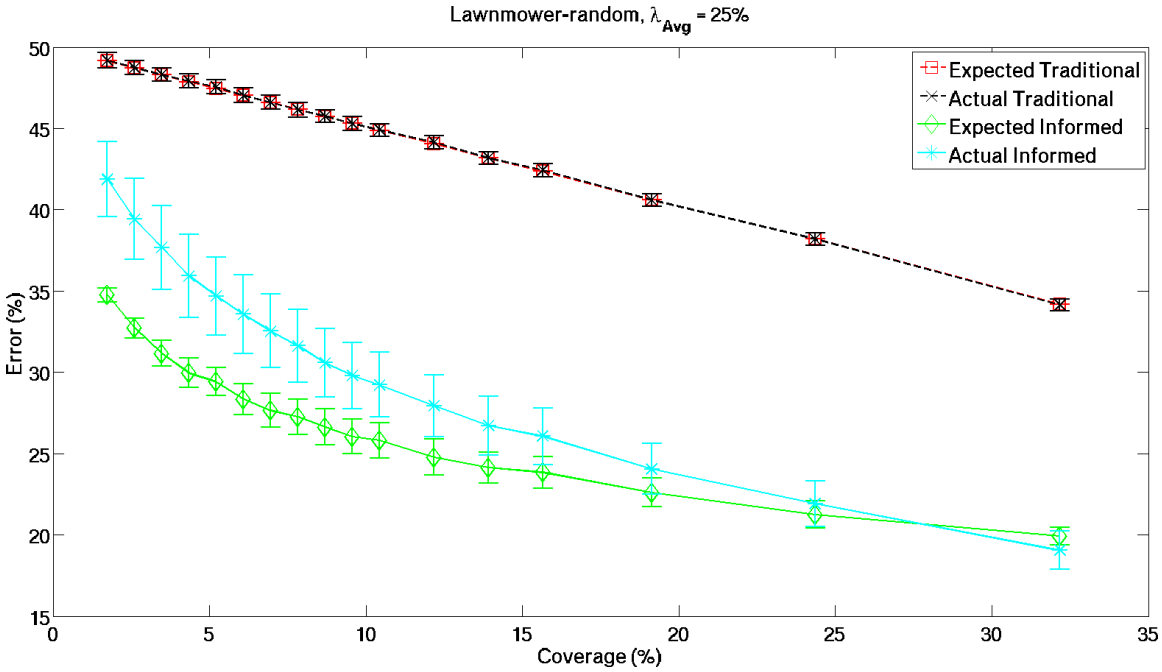


Figure 79: Performance of informed estimation using lawnmower-random navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=25%.

method produces a more linear trend based on the samples collected using the lawnmower-traditional navigation (Figure 77). Although, at this higher spatial frequency, the expected error of the informed method no longer produces an upper bound on the actual

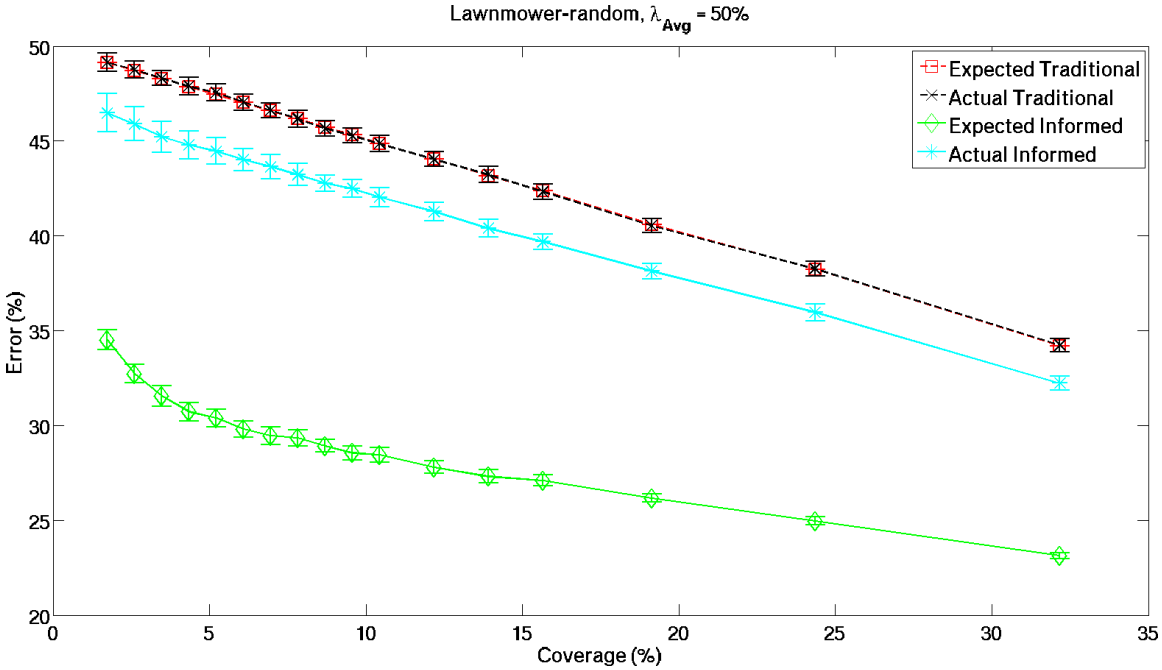


Figure 80: Performance of informed estimation using lawnmower-random navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=50%.

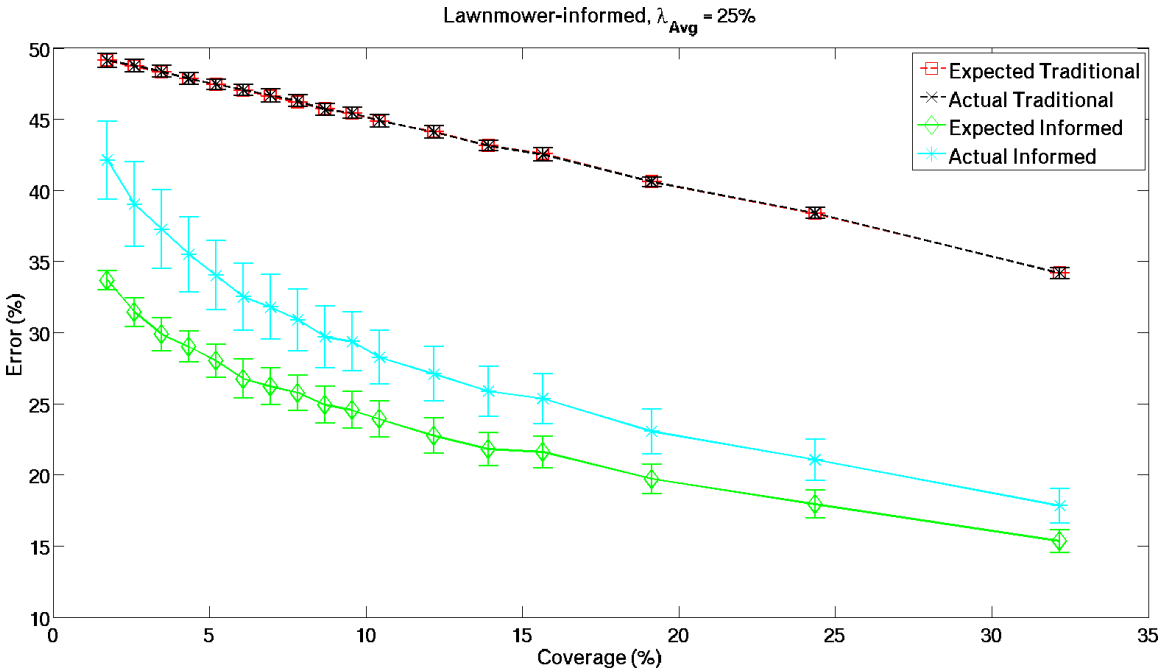


Figure 81: Performance of informed estimation using lawnmower-informed navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=25%.

error, the expected error better approximates the actual error. This is also verified in Figures 79 and 81, where the average expected error more closely approximates

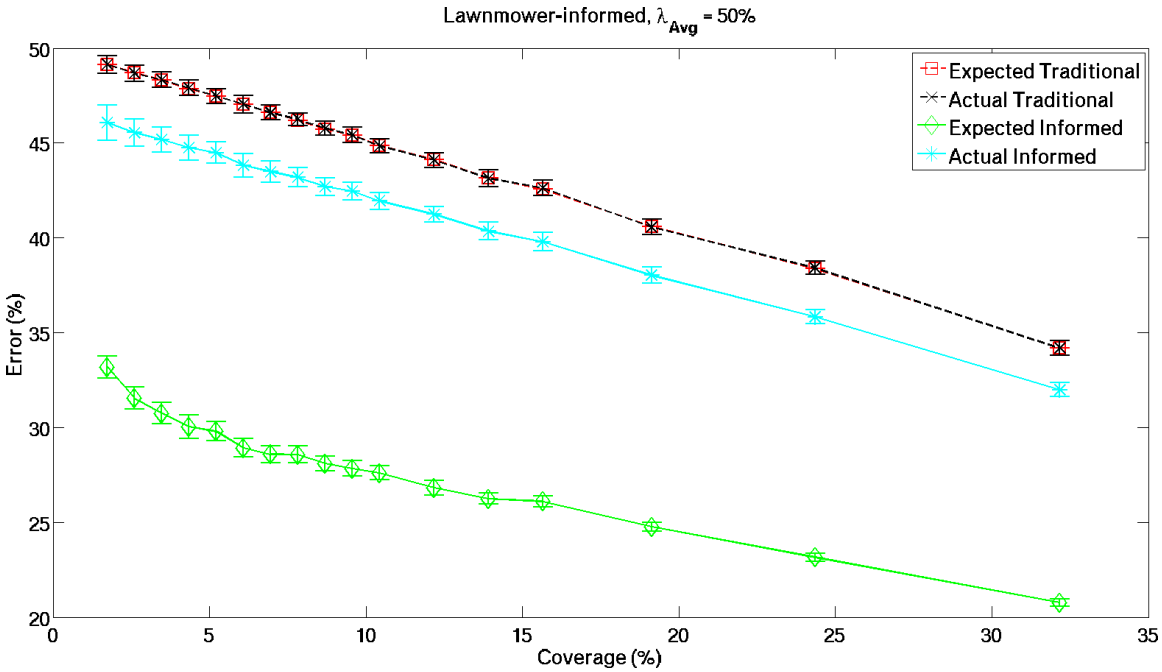


Figure 82: Performance of informed estimation using lawnmower-informed navigation as a sampling scheme as applied to simulated DEM data at λ_{Avg}=50%.

the average actual error for all coverages evaluated. Similarly, the trend of actual error for traditional estimation almost exactly matches the projected expected error. This close match shows the robustness of traditional estimation under changing conditions of spatial frequency, yet, it lacks the improvement in estimation error that our method provides. The improvement of our informed estimation method gives further credence to using nearest-neighbor information to influence estimation decisions for reconstructing a dual-class environment across a range of spatial frequencies.

We acknowledge how, for tested terrains with larger spatial frequencies ($\lambda_{Avg} \geq$ 50%), using traditional estimation to generate expected error is favored over our informed method. Still, with respect to actual error, our process of estimating the class designation (Equation (8)) is more useful for reconstructing more accurate spaces than that of the traditional approach, regardless of a terrain's spatial frequency (Figures 78, 80, and 82).

APPENDIX C

PERFORMANCE AND RESOURCE DATA

C.1 100 Simulated DEMs

Table 14: Relevant performance and resource data for lawnmower-traditional navigation across all coverages for simulated DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.854	32.327	8.584	14.820	0.885	50.9414	0.885
1.709	19.706	6.331	7.962	1.785	36.597	1.785
2.563	12.622	4.178	5.331	2.702	22.957	2.702
3.418	10.607	2.967	5.263	3.569	17.921	3.569
4.272	8.511	2.505	4.219	4.541	16.189	4.541
5.126	7.405	2.115	2.616	5.547	13.316	5.547
5.981	6.040	1.531	2.851	6.238	9.316	6.238
6.836	5.585	1.392	2.790	7.125	8.975	7.125
7.690	5.115	1.354	2.420	8.305	8.567	8.305
8.544	4.596	1.110	2.435	9.054	7.183	9.054
9.399	4.286	0.988	2.246	9.950	6.405	9.950
10.253	3.960	0.962	1.868	11.043	6.193	11.043
11.962	3.449	0.848	1.739	12.404	5.512	12.404
13.671	3.085	0.693	1.512	14.142	4.885	14.142
15.380	2.956	0.702	1.361	16.440	5.074	16.440
18.798	2.412	0.554	1.119	19.612	3.894	19.612
23.924	2.000	0.455	0.938	24.254	3.251	24.254
31.614	1.519	0.347	0.643	31.623	2.548	31.623
47.849	1.133	0.265	0.461	44.721	1.777	44.721

Table 15: Relevant performance and resource data for lawnmower-random navigation across all coverages for simulated DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.854	27.473	4.010	18.064	5.104	36.620	5.104
1.709	17.969	4.945	9.316	7.258	29.353	7.258
2.563	12.425	3.476	5.837	9.068	19.977	9.068
3.418	9.701	2.636	4.839	10.444	15.267	10.444
4.272	7.748	2.206	3.720	11.878	13.701	11.878
5.127	6.475	1.863	2.994	13.177	10.979	13.177
5.981	5.352	1.381	2.722	14.007	8.771	14.007
6.836	4.642	1.243	2.132	15.082	8.015	15.082
7.690	4.136	1.146	1.928	16.389	6.934	16.389
8.544	3.648	0.964	1.694	17.109	6.200	17.109
9.399	3.268	0.863	1.543	18.092	5.596	18.092
10.253	3.111	0.812	1.490	19.028	5.459	19.028
11.962	2.605	0.620	1.316	20.442	4.053	20.442
13.671	2.313	0.567	1.202	21.761	3.599	21.761
15.380	2.197	0.554	1.051	23.944	4.144	23.944
18.798	1.734	0.393	0.900	26.144	2.896	26.144
23.924	1.461	0.319	0.771	29.802	2.367	29.802
31.614	1.253	0.285	0.681	33.765	2.200	33.765
47.849	1.256	0.272	0.612	46.703	1.966	46.703

Table 16: Relevant performance and resource data for lawnmower-informed navigation across all coverages for simulated DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.854	30.315	6.098	12.250	1.252	45.792	1.252
1.709	16.592	5.326	5.898	2.525	29.353	2.887
2.563	10.438	2.917	4.423	3.821	18.442	4.413
3.418	7.430	2.220	3.433	5.295	12.302	5.048
4.272	5.881	1.759	2.155	6.828	10.170	7.206
5.128	4.869	1.400	2.344	7.845	8.333	8.382
5.981	3.962	1.091	1.943	9.142	6.692	9.921
6.836	3.401	0.830	1.671	10.405	5.293	10.922
7.690	3.162	0.892	1.278	11.910	5.323	12.129
8.544	2.765	0.705	1.384	12.804	4.340	13.655
9.399	2.522	0.636	1.074	14.072	3.924	15.308
10.253	2.354	0.621	1.142	15.748	4.023	16.442
11.962	1.903	0.462	0.802	18.033	2.972	18.064
13.671	1.627	0.396	0.711	19936	2.488	21.442
15.380	1.641	0.441	0.809	22.533	3.009	22.999
18.798	1.257	0.319	0.575	28.803	2.238	28.753
23.924	0.968	0.227	0.461	32.350	1.543	34.076
31.614	0.756	0.181	0.371	42.888	1.301	43.632
47.849	0.482	0.120	0.204	45.459	0.870	39.795

C.2 20 Realistic DEMs

Table 17: Relevant performance and resource data for lawnmower-traditional navigation across all coverages for realistic DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.664	8.376	4.547	2.933	0.662	20.594	0.662
1.328	4.488	2.656	1.542	1.351	10.913	1.351
1.992	2.905	1.548	1.079	2.040	5.913	2.040
2.656	2.301	1.245	0.849	2.702	5.255	2.702
3.320	1.891	1.007	0.713	3.446	4.008	3.446
3.984	1.562	0.792	0.630	4.163	3.299	4.163
4.647	1.384	0.703	0.570	4.757	2.984	4.757
5.311	1.247	0.636	0.529	5.547	2.870	5.547
5.975	1.100	0.516	0.480	6.238	2.143	6.238
6.639	1.045	0.518	0.462	7.125	2.142	7.125
7.303	0.970	0.480	0.427	7.670	2.091	7.670
7.967	0.879	0.390	0.397	8.305	1.695	8.305
8.631	0.841	0.375	0.388	9.054	1.637	9.054
9.295	0.772	0.326	0.378	9.950	1.380	9.950
10.623	0.718	0.313	0.346	11.043	1.376	11.043
11.951	0.662	0.284	0.326	12.403	1.343	12.403
13.942	0.575	0.228	0.302	14.142	1.108	14.142
15.934	0.526	0.202	0.276	16.440	0.955	16.440
19.254	0.448	0.160	0.254	19.612	0.808	19.612
24.565	0.391	0.136	0.224	24.254	0.720	24.254

Table 18: Relevant performance and resource data for lawnmower-random navigation across all coverages for realistic DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.664	8.439	3.426	3.849	4.163	15.549	4.162
1.328	3.953	1.546	1.838	6.210	6.733	6.238
1.992	2.615	1.033	1.208	7.669	4.929	7.784
2.656	1.931	0.726	0.893	8.821	3.441	8.821
3.320	1.598	0.576	0.731	10.036	2.709	10.190
3.984	1.360	0.523	0.619	11.027	2.636	11.256
4.647	1.152	0.447	0.485	11.897	2.183	11.897
5.311	1.029	0.408	0.432	12.954	2.021	12.954
5.975	0.910	0.362	0.375	13.801	1.868	13.801
6.639	0.832	0.330	0.353	14.876	1.563	15.231
7.303	0.758	0.301	0.319	15.464	1.530	15.464
7.967	0.705	0.278	0.289	16.214	1.431	16.214
8.631	0.639	0.257	0.280	16.958	1.332	16.958
9.295	0.591	0.228	0.285	17.989	1.010	17.920
10.623	0.550	0.227	0.241	18.964	1.193	18.964
11.951	0.495	0.211	0.235	20.308	1.102	20.308
13.942	0.415	0.160	0.206	21.808	0.838	21.808
15.934	0.380	0.141	0.191	23.913	0.656	23.842
19.254	0.330	0.125	0.158	26.234	0.586	26.169
24.565	0.284	0.106	0.135	29.645	0.550	29.687

Table 19: Relevant performance and resource data for lawnmower-informed navigation across all coverages for realistic DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.664	7.612	2.381	3.190	0.937	13.005	0.937
1.328	3.493	1.383	1.641	1.911	5.909	2.152
1.992	2.214	0.704	1.016	2.886	3.946	3.154
2.656	1.643	0.630	0.753	3.821	2.870	4.106
3.320	1.297	0.503	0.704	4.874	2.334	4.874
3.984	1.126	0.368	0.513	5.887	1.900	5.969
4.647	0.916	0.301	0.407	6.727	1.623	6.727
5.311	0.799	0.296	0.370	7.845	1.629	8.046
5.975	0.718	0.232	0.401	8.822	1.324	8.822
6.639	0.730	0.257	0.396	10.076	1.298	10.076
7.303	0.605	0.221	0.298	10.847	1.178	10.847
7.967	0.579	0.203	0.279	11.744	1.122	11.744
8.631	0.508	0.180	0.247	12.804	1.024	12.804
9.295	0.511	0.186	0.241	14.072	1.035	14.072
10.623	0.409	0.151	0.176	15.617	0.818	15.617
11.951	0.388	0.131	0.201	17.541	0.687	17.541
13.942	0.311	0.098	0.151	20.000	0.540	20.000
15.934	0.300	0.108	0.142	23.250	0.547	23.250
19.254	0.236	0.086	0.103	27.472	0.460	27.672
24.565	0.187	0.059	0.081	33.864	0.291	34.705

C.3 10 MODIS DEMs

Table 20: Relevant performance and resource data for lawnmower-traditional navigation across all coverages for MODIS DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.854	23.115	10.993	14.034	0.885	48.847	0.885
1.709	18.533	6.190	11.871	1.785	30.359	1.785
2.563	18.636	4.600	13.936	2.702	28.212	2.702
3.418	14.350	3.574	10.155	3.569	20.854	3.569
4.272	14.851	3.767	10.926	4.541	21.641	4.541
5.127	13.500	3.151	9.822	5.547	19.115	5.547
5.981	12.060	2.913	9.316	6.238	17.860	6.238
6.836	11.445	2.887	8.083	7.125	17.422	7.125
7.690	10.929	2.362	8.136	8.305	15.274	8.305
8.544	10.206	2.499	6.699	9.054	14.480	9.054
9.399	9.425	1.900	6.601	9.950	12.265	9.950
10.253	9.430	2.052	6.571	11.043	13.202	11.043
11.962	8.033	1.798	5.482	12.403	11.009	12.403
13.671	7.599	1.574	5.172	14.142	9.747	14.142
15.380	7.206	1.534	4.620	16.440	9.346	16.440
18.798	6.174	1.353	4.038	19.612	8.045	19.612
23.924	5.276	1.111	3.312	24.254	6.465	24.254
31.614	4.101	0.864	2.578	31.623	5.006	31.623
47.849	3.130	0.604	1.943	44.721	3.871	44.721

Table 21: Relevant performance and resource data for lawnmower-random navigation across all coverages for MODIS DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.854	21.598	7.495	14.737	5.030	35.440	5.0303
1.709	16.492	6.850	10.336	7.149	32.287	7.1493
2.563	14.318	5.741	9.134	8.985	26.064	8.9851
3.418	12.767	4.366	8.227	10.315	22.064	10.315
4.272	11.673	4.119	7.637	12.032	19.773	12.032
5.127	10.871	3.625	6.904	13.266	18.503	13.266
5.981	10.071	3.141	6.647	14.145	16.802	14.145
6.836	9.054	3.026	6.004	15.020	16.068	15.020
7.690	8.752	2.786	5.263	16.157	14.465	16.157
8.544	8.092	2.764	5.013	16.889	14.079	16.889
9.399	7.771	2.220	5.285	17.830	12.416	17.830
10.253	7.618	2.274	4.915	18.801	11.947	18.801
11.962	6.371	2.009	3.705	20.155	10.352	20.155
13.671	5.949	1.749	3.660	21.554	8.870	21.554
15.380	5.564	1.664	3.478	23.612	8.696	23.612
18.798	4.729	1.295	2.813	26.201	6.715	26.201
23.924	4.282	1.009	2.760	28.342	5.641	28.342
31.614	3.614	0.801	2.329	33.128	4.620	33.128
47.849	3.616	0.777	2.238	24.290	4.529	24.290

Table 22: Relevant performance and resource data for lawnmower-informed navigation across all coverages for MODIS DEMs.

Coverage (%)	Avg. err. (%)	Std. Dev. (%)	Min. err. (%)	Resources (%)	Max. err. (%)	Resources (%)
0.854	19.927	8.340	10.843	1.543	36.227	1.579
1.709	15.564	6.202	9.686	3.054	27.229	2.806
2.563	12.961	4.557	8.204	4.475	21.618	4.522
3.418	12.244	4.521	7.887	5.735	21.611	5.761
4.272	11.326	3.687	7.448	7.080	19.546	7.312
5.127	10.320	3.399	6.647	9.175	17.732	8.314
5.981	9.304	2.957	6.216	9.997	16.106	9.854
6.836	8.395	2.865	4.991	10.908	14.662	11.609
7.690	8.048	2.771	5.134	13.335	13.739	13.349
8.544	7.291	2.452	4.628	14.846	12.439	13.644
9.399	6.802	2.171	4.219	15.708	11.448	15.014
10.253	6.635	2.132	4.053	17.282	11.062	16.466
11.962	5.763	1.851	3.342	20.159	9.399	19.543
13.671	5.253	1.525	3.047	21.599	8.234	22.047
15.380	5.075	1.506	2.858	26.854	7.766	24.981
18.798	4.178	1.142	2.435	29.966	5.966	30.333
23.924	3.460	0.909	2.079	34.096	4.870	35.457
31.614	2.664	0.683	1.474	42.714	3.546	44.494
47.849	2.023	0.530	1.127	41.325	2.730	44.584

C.4 Single Real DEM

Table 23: Relevant performance and resource data for lawnmower-traditional navigation across all coverages for a real DEM.

Coverage (%)	Error (%)	Resources (%)
0.488	38.756	0.503
0.975	12.244	1.010
1.463	6.119	1.515
1.950	8.826	2.040
2.438	5.841	2.563
2.925	3.985	3.029
3.413	5.363	3.569
3.900	4.060	4.163
4.388	3.109	4.541
4.876	3.826	5.256
5.363	2.980	5.547
5.851	2.502	6.238
6.338	3.264	6.652
6.826	2.557	7.125
7.313	2.328	7.670
7.801	2.831	8.305
8.776	2.308	9.054
9.264	2.468	9.950
10.726	2.169	11.043
11.701	1.756	12.403
13.652	1.831	14.142
16.090	1.756	16.440
19.015	1.517	19.612
23.891	1.224	24.254

Table 24: Relevant performance and resource data for lawnmower-random navigation across all coverages for a real DEM.

Coverage (%)	Error (%)	Resources (%)
0.488	32.866	3.828
0.975	12.642	5.462
1.463	6.612	6.737
1.950	5.791	7.777
2.438	4.786	8.721
2.925	3.711	9.557
3.413	3.114	10.214
3.900	2.642	11.235
4.388	2.866	11.752
4.876	2.418	12.551
5.363	2.080	13.006
5.851	1.831	13.980
6.338	1.522	14.318
6.826	1.532	14.864
7.313	1.746	15.328
7.801	1.507	16.066
8.776	1.209	16.964
9.264	1.204	18.020
10.726	0.995	18.844
11.701	0.871	20.268
13.652	1.070	21.662
16.090	0.766	23.529
19.015	0.721	26.269
23.891	0.756	28.704

Table 25: Relevant performance and resource data for lawnmower-informed navigation across all coverages for a real DEM.

Coverage (%)	Error (%)	Resources (%)
0.488	50.269	0.959
0.975	14.692	1.892
1.463	10.771	2.836
1.950	4.299	4.244
2.438	3.134	4.477
2.925	2.388	5.174
3.413	2.413	6.039
3.900	2.159	7.141
4.388	1.612	7.609
4.876	1.373	9.001
5.363	1.617	9.314
5.851	1.129	9.897
6.338	1.423	10.873
6.826	0.940	11.318
7.313	1.035	12.341
7.801	1.095	12.613
8.776	0.776	14.102
9.264	0.871	14.865
10.726	0.726	17.689
11.701	0.786	18.263
13.652	0.587	21.664
16.090	0.453	22.540
19.015	0.448	28.628
23.891	0.353	29.961

REFERENCES

[1] Ayeni, O. O., "Optimum sampling for digital terrain models: A trend towards automation," *Photogrammetric Engineering and Remote Sensing*, vol. 48, no. 11, pp. 1687–1694, 1982.

[2] Samal, A., "Pincock perspectives: Basics of variogram analysis," May 2007.

[3] Low, K. H., Gordon, G. J., Dolan, J. M., and Khosla, P., "Adaptive sampling for multi-robot wide-area prospecting," tech. rep., Carnegie Mellon University, 2005.

[4] Low, K. H., Dolan, J. M., and Khosla, P., "Adaptive multi-robot wide-area exploration and mapping," in *Proceedings of the 7th Int. Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, p. 2330, 2008.

[5] Rahimi, M., Pon, R., Kaiser, W. J., Sukhatme, G. S., Estrin, D., and Srivastava, M., "Adaptive sampling for environmental robotics," in *IEEE Int. Conference on Robotics and Automation*, 2004.

[6] Batalin, M. A., Rahimi, M., Yu, Y., Liu, D., Kansal, A., Sukhatme, G. S., Kaiser, W. J., Hansen, M., Pottie, G. J., Srivastava, M., and Estrin, D., "Call and response: experiments in sampling the environment," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, SenSys '04, pp. 25–38, 2004.

[7] Bourgeois, B. S., Brandon, D. L., Cheramie, J. J., and Gravley, J., "Efficient hydrographic survey planning using an environmentally adaptive approach," in *DoD Technical Report*, (Stennis Space Center, MS), 2006.

[8] *Precise Geodetic Infrastructure: National Requirements for a Shared Resource*. Washington, DC: The National Academies Press, 2010.

[9] NOAA, "http://oceanservice.noaa.gov/."

[10] Stroeve, J., Box, J. E., Gao, F., Liang, S., Nolin, A., and Schaaf, C., "Accuracy assessment of the modis 16-day albedo product for snow: comparisons with greenland in situ measurements," *Remote Sensing of Environment*, vol. 94, pp. 46 – 60, 2005.

[11] Spikes, V. B. and Hamilton, G. S., "Glas calibration-validation sites established on the west antarctic ice sheet," in *International Symposium on Remote Sensing of Environment*, (Honolulu, Hawaii), 2003.

[12] Davis, C. H. and Segura, D. M., "An algorithm for time series analysis of ice sheet surface elevations from satellite altimetry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 202 – 206, 2001.

[13] Keller, K., Casassa, G., Rivera, A., Forsberg, R., and Gundestrup, N., "Airborne laser altimetry survey of glaciar tyndall, patagonia," *Journal of Global and Planetary Change*, vol. 59, pp. 101 – 109, 2007.

[14] Parker, L. T., English, B., Chavis, M. A., and Howard, A. M., "Improvements to satellite-based albedo measurements using in-situ robotic surveying techniques," in *AIAA Infotech@Aerospace Conference*, April 20 – 22 2010.

[15] Williams, S., Parker, L. T., and Howard, A. M., "Calibration and validation of earth-observing sensors using deployable surface-based sensor networks," in *IEEE Journal of Selected Topics in Earth Observations and Remote Sensing*, 2009.

[16] Parker, L. T. and Howard, A. M., "Real-time robotic surveying for unexplored arctic terrain," in *NASA Earth Science Technology Forum*, June 22 – 24 2010.

[17] Space Studies Board, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, 2007.

[18] Wang, J., Haining, R., and Cao, Z., "Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning," *International Journal of Geographical Information Science*, vol. 24, no. 4, pp. 523–543, 2010.

[19] Reilly, C. and Gelman, A., "Weighted classical variogram estimation for data with clustering," *Technometrics*, vol. 49, no. 2, pp. 184–194, 2007.

[20] Marchant, B. P. and Lark, R. M., "Optimized sample schemes for geostatistical surveys," *Mathematical Geology*, vol. 39, no. 1, pp. 113–134, 2007.

[21] Balce, A. E., "Determination of optimum sampling interval in grid digital elevation models (dem) data acquisition," *Photogrammetric Engineering and Remote Sensing*, vol. 53, no. 3, pp. 323–330, 1987.

[22] Billa, M., Cassard, D., Lips, A. L. W., Bouchot, V., Tourliere, B., Stein, G., and Guillou-Frottier, L., "Predicting gold-rich epithermal and porphyry systems in the central andes with a continental-scale metallogenic gis," *Ore Geology Reviews*, vol. 25, pp. 39–67, 2004.

[23] Atkinson, P. M. and Llyod, C. D., "Non-stationary variogram models for geostatistical sampling optimisation: An empirical investigation using elevation data," *Computers & Geosciences*, vol. 33, pp. 1285–1300, 2007.

[24] Webster, R. and Oliver, M. A., *Geostatistics for Environmental Scientists*. John Wiley & Sons Publishing, 2003.

[25] Morrison, J. L., "A link between cartographic theory and mapping practice; the nearest neighbor statistic," *Geographical Review*, vol. 60, 1970.

[26] Li, J. and Heap, A. D., *A Review of Spatial Interpolation Methods for Environmental Scientists*. Australian Government, 2008.

[27] Gneiting, T., Sasvari, Z., and Schlather, M., "Analogies and correspondences between variograms and covariance functions," Tech. Rep. 056, National Research Center for Statistics and the Environment, October 2000.

[28] Wingham, D. J., Shepherd, A., Muir, A., and Marshall, G. J., "Mass balance of the antarctic ice sheet," *Philosophical Transactions of the Royal Society A*, vol. 364, pp. 1627 – 1635, 2006.

[29] Keutterling, A. and Thomas, A., "Monitoring glacier elevation and volume changes with digital photogrammetry and gis at gepatschferner glacier, austria," *International Journal of Remote Sensing*, vol. 27, pp. 4371 – 4380, October 2006.

[30] Dinkins, C. P. and Jones, C., *Soil Sampling Strategies*. Montana State University, 2008.

[31] Meersmans, J., Ridder, F. D., Canters, F., Baets, S. D., and Molle, M. V., "A multiple regression approach to assess the spatial distribution of soil organic carbon (soc) at the regional scale (flanders, belgium)," *Geoderoma*, no. 143, pp. 1–13, 2008.

[32] Carranza, E. J. and Sadeghi, M., "Predictive mapping of prospectivity and quantitative estimation of undiscovered vms deposits in skellefte district (sweden)," *Ore Geology Reviews*, no. 38, pp. 219–241, 2010.

[33] Carranza, E. J., "Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity," *Computers & Geosciences*, no. 35, pp. 2032–2046, 2009.

[34] Tunstel, E., Anderson, G. T., and Wilson, E. E., "Autonomous mobile surveying for science rovers using in situ distributed remote sensing," in *IEEE Int. Conference on Systems, Man, and Cybernetics*, (Pasedena, CA), 2007.

[35] Tunstel, E., Dolan, J., Fong, T. W., and Schreckenghost, D., "Mobile robotic surveying performance for planetary surface site characterization," in *Performance Evaluation and Benchmarking of Intelligent Systems* (Tunstel, E. and Messina, E., eds.), Springer, August 2009.

[36] Tunstel, E., McFarlane, L., and Mejia, V., "Rover prototype for mobile surveying technology development," in *NSBE Aerospace Conference*, (Los Angeles, CA), 2010.

[37] Anderson, M. O., Kinoshita, R. A., McKay, M. D., Willis, W. D., Gunderson, R. W., and Flann, N. S., "Mobile robotic teams applied to precision agriculture," in *8th International Topical Meeting on Robotics and Remote Systems*, (Idaho Falls, ID), 1999.

[38] Fong, T., Allan, M., Bouyssounouse, X., Bualat, M. G., Deans, M. C., Edwards, L., Fluckiger, L., Keely, L., Lee, S. Y., Lees, D., To, V., and Utz, H., "Robotic site survey at haughton crater," in *9th Int. Symposium on Artificial Intelligence, Robotics and Automation in Space*, (Los Angeles, CA), 2008.

[39] Fong, T. W., Bualat, M., Edwards, L., Flueckiger, L., Kunz, C., Lee, S. Y., Park, E., To, V., Utz, H., Ackner, N., Armstrong-Crews, N., and Gannon, J., "Human-robot site survey and sampling for space exploration," in *AIAA Space 2006*, September 1996.

[40] Low, K. H., Gordon, G. J., Dolan, J. M., and Khosla, P., "Adaptive sampling for multi-robot wide-area exploration," in *IEEE Int. Conference on Robotics and Automation*, pp. 755–760, 2007.

[41] NASA, "http://lunarscience.nasa.gov/robots/2010/robotic-tech."

[42] Jin, J., *Optimal field coverage path planning on 2D and 3D surfaces.* PhD dissertation, Iowa State University, Department of Agricultural and Biosystems Engineering, 2009.

[43] Hashemi, R. R., Jin, L., Anderson, G. T., Wilson, E., and Clark, M. R., "A comparison of search patterns for cooperative robots operating in remote environment," in *IEEE Int. Conference on Information Technology: Coding and Computing*, (Las Vegas, NV), 2001.

[44] Hashemi, R. R., Jin, L., Jones, S., Owens, D., and Anderson, G., "A rule-based system for localization of water on the surface of mars," in *IEEE Int. Conference on Information Technology: Coding and Computing*, (Las Vegas, NV), 2001.

[45] Durrant-Whyte, H. and Bailey, T., "Simultaneous localization and mapping: part i," *Robotics Automation Magazine, IEEE*, vol. 13, pp. 99 –110, june 2006.

[46] Dantu, K., Rahimi, M., Shah, H., Babel, E., Dhariwal, A., and Sukhatme, G. S., "Robomote: Enabling mobility in sensor networks," in *In IEEE/ACM Fourth International Conference on Information Processing in Sensor Networks (IPSN-SPOTS*, pp. 404–409, 2005.

[47] Cortes, J., "Distributed kriged kalman filter for spatial estimation," *IEEE Transactions on Automatic Control*, December 2009.

[48] Castello, C. C., Davari, A., and Chen, R., "Optimal sensor placement strategy for environmental monitoring using wireless sensor networks," in *In Proceedings of the 42nd South Eastern Symposium on System Theory*, (Tyler, TX), March 2010.

[49] Cover, T. and Hart, P., "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on*, vol. 13, pp. 21 –27, january 1967.

[50] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3rd ed., 1991.

[51] Clark, P. J. and Evans, F. C., "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, pp. 445 –453, October 1954.

[52] Liu, L., Crowe, T. G., Roberge, M., and Bakambu, J. N., "Vision-based exploration algorithms for rough terrain modeling using triangular mesh maps," in *IEEE Int. Workshop of Robotic Sensors and Environments*, (Ottawa, Canada), 2007.

[53] Rahimi, M., Hansen, M., Kaiser, W. J., Sukhatme, G. S., and Estrin, D., "Adaptive sampling for environmental robotics," in *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, (New Orleans, LA), pp. 3692 – 3698, 2005.

[54] Heavner, M. L., Fatland, D. R., Hood, E. W., and Connor, C., "A cryospheric sensor web use case on a small temperate glacier," in *Proceedings of the Earth Science Technology Conference*, (Washington, DC), June 2008.

[55] Moghaddam, M., Burgin, M., Castillo, A., Entekhabi, D., Goykhman, Y., Kakhbod, A., Li, K., Liu, M., Nayyar, A., Silva, A., Teneketzis, D., Wang, Q., and Wu, X., "Soil moisture sensing controller and optimal estimator (soilscape): First deployment of the wireless sensor network and latest progress on soil moisture satellite retrieval validation strategies," in *NASA Earth Science Technology Forum*, June 21 – 23 2011.

[56] Parker, L. T. and Howard, A. M., "Adaptive robot navigation protocol for estimating variable terrain elevation data," in *IEEE Int. Conference on Systems, Man, and Cybernetics*, (Anchorage, AK), October 2011.

[57] Abdullah, Q. A., "Mapping matters: The layman's perspective on technical theory and practical applications of mapping and gis," *Journal of Photogrammetric Engineering & Remote Sensing*, vol. 74, pp. 683 – 685, June 2008.

[58] Subcommittee for Base Cartographic Data, Fedreal Geographic Data Committee, *Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy*, fgdc-std-007.3-1998 ed., 1998.

[59] Greenwalt, C. R. and Schultz, M. E., "Principles of error theory and cartographic applications," Tech. Rep. 96, Aeronautical Chart & Information Center (ACIC) Technical Report, February 1962.

[60] Fournier, A., Fussell, D., and Carpenter, L., "Computer rendering of stochastic models," in *Communications of the ACM*, June 1982.

[61] Roberts, L. G., "Machine perception of three-dimensional solids," in *Lincoln Laboratory Technical Report 513*, 1963.

[62] Gonzalez, R. C. and Woods, R. E., *Digital Image Processing*. Prentice Hall, 2002.

[63] Pratt, W. K., *Digital Image Processing*. Wiley, 1978.

[64] Canny, J., "A computational approach to edge detection," in *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–714, 1986.

[65] de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M., *Computational Geometry: Algorithms and Applications*. Springer-Verlag, 2008.

[66] Sandwell, D. T., "Biharmonic spline interpolation of geos3 and seasat altimeter data," *Geophysical Research Letters*, vol. 14, 1987.

[67] Barsky, B. and DeRose, T., "Geometric continuity of parametric curves: three equivalent characterizations," *Computer Graphics and Applications, IEEE*, vol. 9, pp. 60 –69, nov 1989.

[68] Mei, H., Parker, L. T., and Howard, A. M., "Conversion of gis contour maps into surface digital elevation models," in *IEEE Int. Conference on Systems, Man, and Cybernetics*, (Anchorage, AK), October 2011.

[69] Shuman, D., Nayyar, A., Mahajan, A., Goykhman, Y., Li, K., Liu, M., Teneketzis, D., Moghaddam, M., and Entekhabi, D., "Measurement scheduling for soil moisture sensing: From physical models to optimal control," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1918 –1933, 2010.